Archive of SID

1st International Conference on new achievements in Science & Technology
24th August 2017
اولین کنفرانس بین المللی دستاوردهای نوین در علوم و تکنولوژی
۲ شهریور ماه ۱۳۹۶ – تهران

داده‌کاوی هیبریدی و تجزیه و تحلیل خوشه‌ای یادگیری ماشین برای پیش‌بینی نرخ جایگزینی متخصصان حوزه فناوری

سعید آرامش۱، وحید محمدیان۲*، هادی بارانی برواتی۳، مینا ضیایی۴

۱ گروه کامپیوتر دانشکده فنی مهندسی، دانشگاه آزاد اسلامی، واحد ایرانشهر، سیستان و بلوچستان، ایران
آدرس پست الکترونیک: Saeed.aramesh1997@gmail.com
۲ مدرس دانشگاه آزاد اسلامی، واحد ایرانشهر، سیستان و بلوچستان، ایران
آدرس پست الکترونیک: vmcuni@gmail.com
۳ عضو هیئت علمی دانشگاه آزاد اسلامی، واحد ایرانشهر، سیستان و بلوچستان، ایران
آدرس پست الکترونیک: Hb_200@yahoo.com
۴ مدرس دانشگاه آزاد اسلامی، واحد ایرانشهر، سیستان و بلوچستان، ایران
آدرس پست الکترونیک: mina.ziaei89@gmail.com

چکیده

این مطالعه تجزیه و تحلیل خوشه ای برای داده کاوی و یادگیری ماشین را برای پیش بینی روند تکنیک هایی شامل شبکه عصبی مصنوعی هیبرید و تحلیل خوشه‌ای که تحت نقشه خود سازماندهی نیز شناخته می‌شوندرا شامل میشود. از این روش خوشه بندی هیبرید برای مطالعه مشخصات فردی روند خوشه‌های جایگزینی استفاده می‌شد. روند جایگزینی متخصصان حوزه تکنولوژی در شرکت‌های مطرح تایوانی مورد بررسی قرار گرفت. نتایج حاصل از این مطالعه نشان می‌دهد که جایگزینی کارکنان بیشتر به خاطر دلایلی چون عدم وجود وفاداری، رهبری و مدیریت مناسب صورت می‌گرفت. با استفاده از اعتبار سنجی متقابل مشخص شد که دقت خوشه بندی برابر با ۹۲,۷٪ است. این مطالعه به بررسی مشکلات مرتبط با از دست دادن سریع منابع انسانی کلیدی می‌پردازد و می‌تواند به سازمان‌ها در بهبود رقابت-پذیری و بازدهی کمک کند.

کلمات کلیدی: روند جایگزینی کارکنان، تحلیل خوشه‌ای، نقشه خود سازماندهی، خوشه‌بندی شبکه عصبی

www.SID.ir

Archive of SID

1st International Conference on new achievements in Science & Technology
24th August 2017
اولین کنفرانس بین المللی دستاوردهای نوین در علوم و تکنولوژی
۲ شهریور ماه ۱۳۹۶ – تهران

data mining and machine learning clustering analysis to predict the turnover rate for technology professionals

Saeed aramesh [1], vahid mohammadian [2], hadi barani barvati [3], mina ziaei [2]

1- Department of Computer, Azad University, Iranshahr, Sistan and Balocheshtan, Iran
Email: *Saeed.aramesh1997@gmail.com*
2- Department of Computer, Azad University, Iranshahr, Sistan and Balocheshtan, Iran
Email: *vmcuni@gmail.com*
۳- Department of Computer, Azad University, Iranshahr, Sistan and Balocheshtan, Iran
Email: *Hb_200@yahoo.com*
4- Department of Computer, Azad University, Iranshahr, Sistan and Balocheshtan, Iran
Email: *mina.ziaei89@gmail.com*

Abstract

This study applies clustering analysis for data mining and machine learning to predict trends in technology professional turnover rates, including the hybrid artificial neural network and clustering analysis known as the self-organizing map (SOM). This hybrid clustering method was used to study the individual characteristics of turnover trend clusters. Using a transaction questionnaire, we studied the period of peak turnover, which occurs after the Chinese New Year, for individuals divided into various age groups. The turnover trend of technology professionals was examined in well-known Taiwanese companies. The results indicate that the high outstanding turnover trend circle was primarily caused by a lack of inner fidelity identification, leadership and management. Based on cross-verification, the clustering accuracy rate was 92.7%. This study addressed problems related to the rapid loss of key human resources and should help organizations learn how to enhance competitiveness and efficiency.

Keywords*: Turnover trend, Clustering analysis, Self-organizing map, Neural network clustering*

www.SID.ir

Archive of SID

مقدمه

بهبود بازدهی منابع انسانی در سازمان همواری یکی از مباحث تحقیقاتی اصلی بوده است. هدف از این مطالعه بررسی عدم توانایی شرکت‌های حوزه فناوری تایوان در حفظ کارکنان بااستعداد و یافتن راهکاری برای عدم اخراج کارکنان و حفظ رقابت‌پذیری شرکت‌ها است. از این رو از یک تکنیک گسسته برای بررسی تکامل شرکت‌ها استفاده کردیم. تمرکز این مطالعه برروی کارکنان با سنین ۲۰ تا ۳۹ سال بود که سهم بالایی از کارکنان جایگزین شده را تشکیل می‌دهند. برای بررسی روند جایگزینی کارکنان از یک مقیاس سنجش مبتنی بر داده‌های واقعی و داخلی استفاده کردیم. از آزمون‌های روایی و پایایی برای اطمینان از به هم پیوسته بودن داده‌ها استفاده کردیم. سایر نتایج را می‌توان با بکارگیری تکنیک‌های محاسبه‌ای در تحلیل خوشه بندی (یا همان نقشه خود سازمانده (SOM) به همراه شبکه عصبی انتشار بازگشتی (BPN)[1] برای داده کاوی حاصل کرد. انتظار می‌رود که SOM-BPN، که می‌توان از آن برای فهم ویژگی‌های مرتبط با خوشه‌های روند جایگزینی استفاده کرد، اطلاعات قابل اطمینانی جهت حمایت از تصمیمات سیسات‌گذاران دراختیار ما قرار دهد.

به منظور تعیین میزان دقت خوشهبندی در مدل SOM-BPN، متغیرها با هم ادغام و انتخاب شدند. این مقاله با ایجاد متغیرهای معتبر و قابل اطمینان بر مبنای اطلاعات بدست آمده از آثار قبلی و متخصصان این حوزه اقدام به پیش بینی تغییرات آینده می‌کند. افزون بر این، این مطالعه از رویکردی مبتنی بر تحلیل خوشه‌ای، SOM و خوشه بندی شبکه عصبی برای تعیین میزان دقت استفاده می‌کند.

۱. بیان مساله و تعاریف

محققان در گذشته فرآیند جایگزینی را به دو دسته تقسیم می‌کردند: «جایگزینی اختیاری» که می‌تواند فردی و جمعی باشد و «جایگزینی اجباری» که می‌تواند به دلایلی مثل بازنشستگی، مرگ، استخدام نامناسب و یا ادغام باشد.

دالتون، تودور و کرکهاردت (۱۹۸۲) جایگزینی را بر حسب وظایف سازمان دسته بندی می‌کنند. بر این اساس جایگزینی اختیاری خود به دو دسته جایگزینی کارکردی و جایگزینی غیر کارکردی تقسیم می‌شود (جدول ۱). اکثر مطالعات سازمانی بیشتر به جایگزینی غیر کارکردی توجه می‌کنند. در مطالعه فعلی، بر جایگزینی اختیاری تاکید داریم. متغیرهای غیرقابل کنترل و بنگاه‌های دارای عملکرد ضعیف در این مطالعه مورد بررسی قرار نگرفته‌اند.

بالا بودن نرخ جایگزینی و تعداد درخواست‌های استعفا تاثیر بسیاری بر عملکرد سازمان می‌گذارد. نیومن (۱۹۷۴)، کراوت (۱۹۷۵)، مبلی (۱۹۷۷)، موبلی، هورنر و هولینگسوورت (۱۹۷۸)، موبلی، گریفت، هند و مگلینو (۱۹۷۹)، میلر (۱۹۷۹) و مایکلز و اسپکتر (۱۹۸۲) همگی بهترین متغیر برای پیش بینی جایگزینی را اصطلاحا «گرایش به جایگزینی» می‌دانند. افزون بر این پورتر و استیرس (۱۹۷۳) نشان دادند که گرایش به جایگزینی یک پدیده رفتاری بوده که به هنگام تجربه شرایط نا رضایتبخش توسط کارمند به وجود می‌آید. موبلی (۱۹۷۷) نشان داد که گرایشات فردی در جایگزینی یکی از عوامل تعیین کننده در این رفتار است. با وجود این، تحقیقات موبلی و همکاران او در سال ۱۹۷۸ نشان می‌دهد که گرایش به جایگزینی حاصل عدم رضایت از کار، نوع نگرش به جایگزینی، یافت مشاغل بهتر و دسترسی به سایر مشاغل است که خود حاکی از آن است که هدف مورد پیش بینی در رفتار جایگزینی را می‌توان از طریق مطالعه روند جایگزینی تعیین کرد.

مطالعه فعلی بر بالا بودن نرخ جایگزینی از سال ۲۰۰۹ در تایوان تمرکز دارد. این مطالعه با کمک متغیرها و توضیحات سایر مطالعات این حوزه در پنج سال گذشته انجام شده است.

---

[1] Back propagation network

www.SID.ir

جدول ۱

طبقه بندی جایگزینی

ارزیابی سازمانی برروی کارکنان		
	عالی	بسیار بد
ارزیابی فردی ازیک سازمان		
جایگزینی اجباری	ماندن در سازمان	اشتباه در استخدام افراد
جایگزینی اختیاری	ترک سازمان توسط کارکنان (جایگزینی غیرکارکردی)	ترک سازمان توسط کارکنان (جایگزینی کارکردی)

۱-۱داده کاوی و روش تحلیل خوشه‌ای

تحلیل‌های قبلی در زمینه گرایش به جایگزینی یا سایر مباحث تحقیقاتی مرتبط بیشتر از آمار توصیفی به همراه تحلیل همبستگی، تحلیل متغیرها، تحلیل واریانس و تحلیل رگرسیون، تحلیل رگرسیون با چند سلسله مراتب و تحلیل ساختاری با مدلسازی معادله استفاده می‌کردند. ساموئل و چیپونزا (۲۰۰۹)، هائو، جونگ و ینهوی (۲۰۰۹) و برخی دیگر از پژوهشگران از شبکه‌های انتشار بازگشتی (BPNها)، رگرسیون لجستیک و تحلیل داده‌های سبد بازار برای پیش بینی روند جایگزینی استفاده کرده‌اند ولی این دسته از تحلیلات ایرادات متعددی دارد. شبکه‌های انتشار بازگشتی بر پایگاه‌های داده‌ای داخلی یک شرکت که از اطلاعات فرم‌های جایگزینی بدست می‌آیند تمرکز دارند. این فرم‌ها حاوی نظرات کارکنان شرکت هستند. اکثر آنها تمایلی به ارائه دلایل اصلی استعفای خود ندارند که خود اعتبار BPN را سوال برانگیز می‌سازد. افزون بر این، سنجش اعتبار رگرسیون لجستیک برای پیش‌بینی امکان‌پذیر نیست. اگرچه ایده و نتایج تحقیقاتی حاصل از تحلیل داده‌های سبد بازار قابل پذیرش هستند، ولی موارد مشابه برای مقایسه وجود ندارد. از این رو تحلیل داده‌های سبد بازار نمی‌تواند کمک چندانی به پیش بینی روند جایگزینی کند.

از آنجا که روش مناسبی برای تحلیل وجود نداشت، از تکنیک داده‌کاوی برای تحلیل خوشه‌ای استفاده کردیم. مطالعه فعلی از RMSE[۲] بهره می‌گیرد که بکارگیری آن از طریق SOM برای ترکیب خوشه بندی دو مرحله ای مبتنی بر تکنیک طبقه‌بندی BPN[۳] در تحلیل خوشه‌ای صورت می‌گیرد.

۲.روش شناسی

این مطالعه در ابتدا به بررسی مجموعه داده‌های موجود در پرسشنامه پرداخت. این پرسشنامه مبتنی بر تعریف هر مرحله و ابزارهای سنجش مورد استفاده برای تعیین محتوای سوالات بود. آزمایش پرسشنامه با استفاده از رویکرد سطح مقطع انجام گرفت و روند آن در شکل ۱ نشان داده شده است.

۲,۱ پیش پردازش داده‌ها

این قسمت دو بخش از مطالعه را بررسی می‌کند: جمع‌آوری داده و ایجاد مدل. بخش دوم شامل بررسی و انتخاب پرسشنامه مناسب می‌شود. از دو فرآیند سیستماتیک برای تبدیل داده‌های اولیه به داده‌های تحقیقاتی مفید و موثر استفاده کردیم.

---

[2] Root mean square error
[3] Back propagation network

www.SID.ir

Archive of SID

۲,۱,۱ تعریف مدل تحقیقاتی براساس روند جایگزینی

در این قسمت، فرض می‌کنیم که پنج عمل مرتبط با روند جایگزینی وجود دارد. یک نمودار علت و معلول برای هر بخش اداری در شکل ۲ نشان داده شده است. در این شکل، فشار کاری و سیاست‌های سازمانی بر رضایت شغلی اثر می‌گذارند و رضایت شغلی و وعده‌های رئیس بر رفتار سازمانی اثر می‌گذارد. افزون بر این، رفتار سازمانی بر روند جایگزینی تاثیر گذار است. نمودار گردشی (فلوچارت) به همراه جزئیات در شکل ۲ آمده است.

۲,۱,۲ نظرسنجی و تفکیک پرسشنامه‌های معتبر

نظرسنجی با استفاده از پرسشنامه در این مطالعه به هنگام فصل جایگزینی بعد از سال نوی چینی صورت گرفته است (یعنی زمانی که بیشترین جایگزینی انجام می‌شود). جامعه آماری هدف در این مطالعه بالاترین گرایش به جایگزینی را داشتند (یعنی افراد با ۲۰ تا ۳۹ سال سن). جامعه آماری را به افراد دارای نوسانات شغلی بالا و پایین تقسیم می‌کنیم. این نظرسنجی براساس نمونه برداری تصادفی سلسله مراتبی و شاخص‌های ارزیابی روایی و پایایی انجام شده است.

سنجش روایی (قابلیت اطمینان) این مطالعه با استفاده از ضریب آلفایکرونباخ (α) صورت گرفت و این مقدار باید بیش از ۰,۷ بدست می‌آمد. به منظور ارزیابی پایایی، اطلاعات اصلی را با استفاده از تحلیل عاملی برای جایگزینی با متغیرهای اولیه و پرهیز از هم خطی استخراج کردیم. سپس متغیرها را براساس قابلیت اطمینان، بار عالمی بزرگتر از ۰,۵ و واریانس توضیحی انتخاب کردیم و آنها را با مطالعات قبلی برای سنجش اعتبار مطالعه فعلی مقایسه کردیم.

۳,۱,۲,۱. طراحی پرسشنامه.

مطالعه فعلی از شش ابزار برای سنجش استفاده می‌کند: (۱) پرسشنامه رضایت مینوستا که با هدف سنجش رضایت شغلی پاسخ دهنده طراحی شده است؛ (۲) مقیاس تعهد سازمانی که براساس یک مقیاس محلی طراحی شده توسط موسسه مدیریت کسب و کار بوده و برای سنجش تعهد داخلی در سازمان بکار می‌رود؛ (۳) مقیاس تعهد ناظر (سوپروایزر) که براساس یک مقیاس محلی طراحی شده توسط موسسه مدیریت کسب و کار بوده و برای سنجش میزان وفاداری داخلی ناظر بکار می‌رود؛ (۴)مقیاس استرس شغلی؛ (۵) مقیاس گرایش به جایگزینی و (۶) مقیاس ادراک سیاست‌های سازمانی. این مطالعه ۲۴ متغیر را با استفاده از تحلیل عاملی استخراج کرده است و این داده‌ها مطابق ۴ متغیر جمعیت شناختی نرمال سازی شده اند. متغیرهای انتخابی در شکل ۳ نشان داده شده‌اند.

۲,۲ تحلیل خوشه‌ای نمونه

این مطالعه نمونه های جمع‌آوری شده را بر حسب نوسانات شغلی در دو دسته بالا و پایین قرار می‌دهد و سپس داده‌ها را با استفاده از یک SOM به همراه شبکه‌های عصبی مصنوعی خوشه بندی می‌کند. در این مطالعه تعداد مناسب خوشه‌ها با استفاده از سه شاخص تعیین می‌شوند.

(۱) یک متغیر عملی که واریماکس را با کمترین تعداد خوشه توضیح می‌دهد.
(۲) تعداد چرخه‌های یادگیری
(۳) خطای مقدار موثر (RMSE[۴])

هدف از این مطالعه یافت مشخصات انحصاری خوشه‌های با بالاترین گرایش به جایگزینی از متغیرهای مختلف است. برای بررسی تاثیر خوشه بندی، این پژوهش از SOM+BPN استفاده می‌کند که اغلب شامل خوشه بندی سلسله مراتبی و غیر سلسله مراتبی

[۴] Root mean square error

www.SID.ir

می‌شود. این مطالعه به سنجش بیشتر اعتبار تفاوت‌های میان متغیرهای حاصل از طبقه بندی ANN می‌پردازد. در ادامه خلاصه‌ای از مراحل اصلی این تحلیل آمده است.

۲,۲,۱ فرآیند خوشه بندی SOM

در این مطالعه از نرم افزار Statistical 7.0 برای بکارگیری ابزارهای اجرایی SOM استفاده شده است. مراحل زیر جزئیات این فرآیند را نشان می‌دهد:

گام ۱: مقدار دهی اولیه به هر سلول عصبی برای نمایش وزن تصمیم جایگزینی، $w_i = [w_{i1}, w_{i2}, w_{i3}, \ldots, w_{ij}]^T \in j$. در این مطالعه، مقدار دهی اولیه به سلول‌های عصبی با استفاده از نمونه‌های استخراجی از مجموعه داده‌های ورودی صورت گرفت.

گام ۲: ارائه یک الگو برای ورودی، $x = [x_1, x_2, x_3, \ldots, x_j]^T \in j$. در تحقیق فعلی، الگوی ورودی مجموعه ای از متغیرهایی بود که نمایانگر فاصله تا تصمیم جایگزینی جدید برای الگوی x و وزن هریک از سلول‌ها (w_i) بودند. همچنین خواهان شناسایی مطابق پذیرترین واحد (w_c) نیز بودیم.

$$\|x - w_c\| = \min\{\|x - w_i\|\} \qquad (1)$$

گام ۳: تعدیل ضرایب سلول مغزی برنده (c) و سایر واحدهای مجاور.

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)], \qquad (2)$$

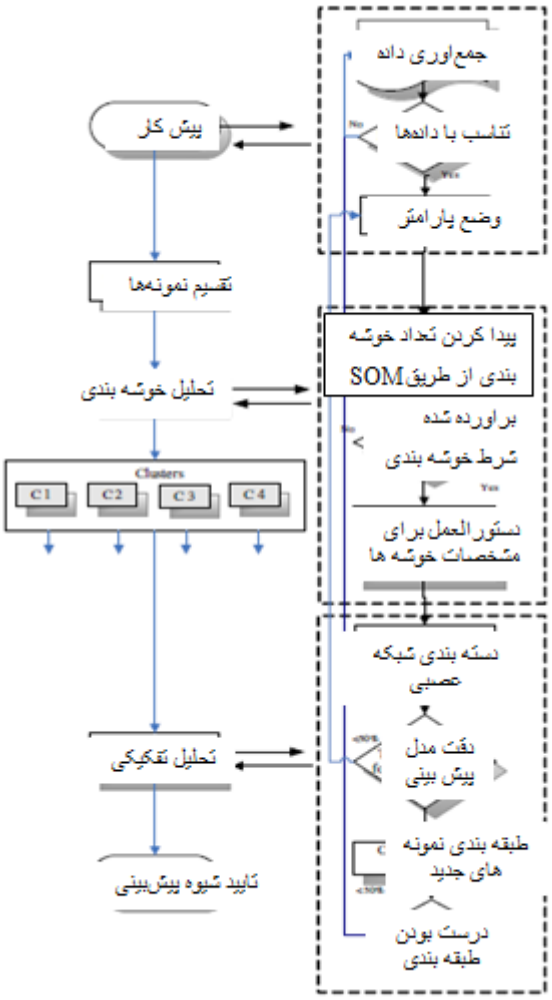
که در آن w_j شاخص نرون مجاور و t عددی صحیح برای نمایش زمان گسسته است. هسته مجاور $h_{ci}(t)$ تابعی از زمان است و فاصله میان نرون مجاور i و نرونبرنده $ch_{ci}(t)$ ناحیه اثر الگوی ورودی بر SOM را مشخص می‌کند. هسته مجاور در SOM از دوقسمت تشکیل می‌شود، تابع همسایگی $h(\|\cdot\|, t)$ و تابع نرخ یادگیری $a(t)$ که در رابطه (۳) نشان داده شده است.

$$h_{ci}(t) = h(\|r_c - r\|_i, t)a(t). \qquad (3)$$

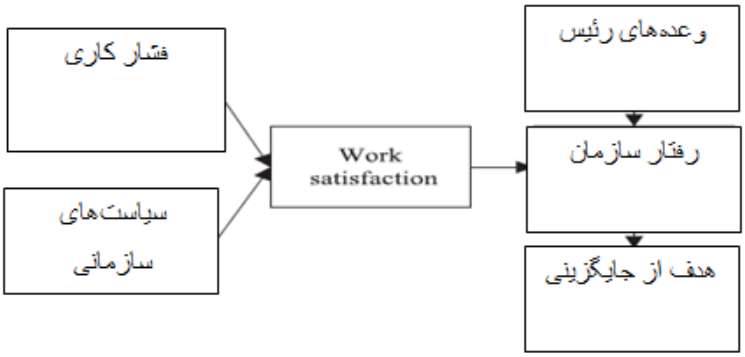
در رابطه r،۳ مکان نرون برروی نقشه دو بعدی است. در این اثر، از تابع همسایگی گاوس استفاده کرده ایم. تابع نرخ یادگیری $a(t)$ تابعی نزولی از زمان است. شکل نهایی هسته مجاورت تحت تابع گاوس از معادله زیر تعریف می‌شود:

www.SID.ir

Archive of SID



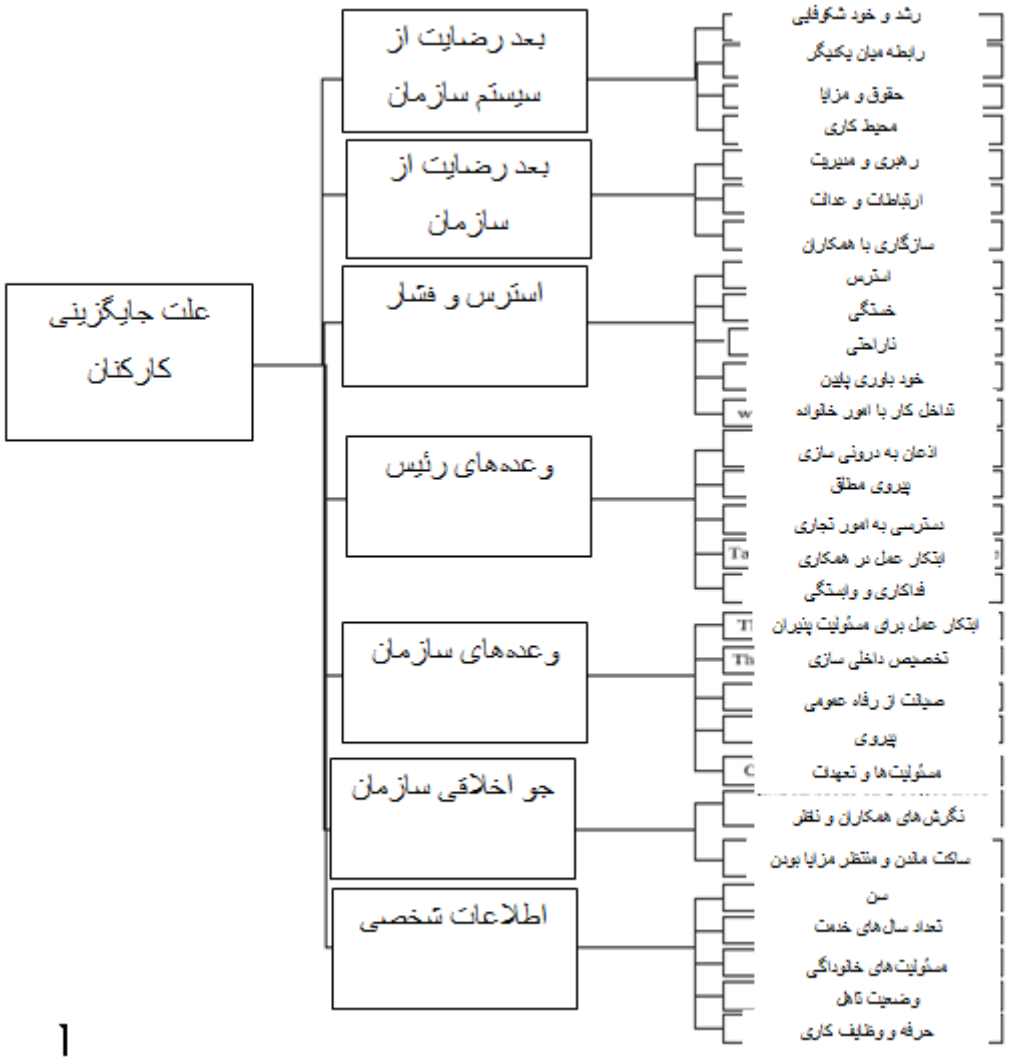
شکل ۱ نمودار گردشی در مطالعه فعلی



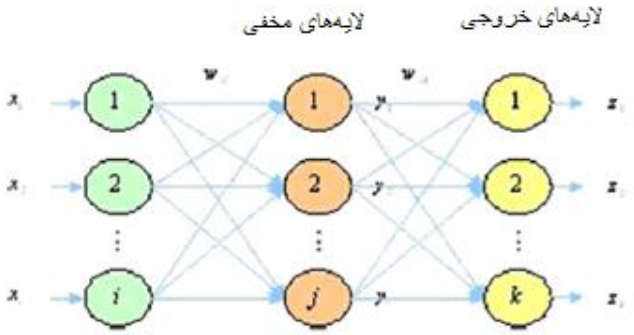
www.SID.ir

Archive of SID

شکل ۲ تعریف مدل تحقیقاتی مطابق گرایش به جایگزینی



شکل ۳ نمودار گردشی در طراحی پرسشنامه



www.SID.ir

شکل ۴ ساختار شبکه عصبی انتشار بازگشتی

$$h_{ci}(t) = \exp\left(\frac{\|r_c - r_i\|}{2\sigma^2(t)}\right)a(t) \qquad (4)$$

که در آن $a(t)$ نشان دهنده عرض هسته است.

گام ۴: مراحل ۲ و ۳ را آنقدر تکرار کنید تا معیار همگرایی برآورده شود. این پژوهش از روش SOM از طریق یک فرآیند یادگیری SOM برای قضاوت در مورد بهترین شیوه خوشه بندی استفاده می‌کند.

۲,۲,۲ روش خوشه بندی شبکه عصبی

شبکه‌های عصبی مصنوعی یکی از سیستم های یادگیری ماشینی هستند. نشان داده شده است که این دسته از شبکه‌ها در تقریب توابع غیرخطی مختلط کارایی دارند. در طبقه بندی، توابع غیرخطی مختلط نمایانگر شکل جداساز میان طبقات است.

از گذشته شبکه‌های عصبی مصنوعی از تعدادی گره در تعداد زیادی لایه تشکیل می‌شوند. داده‌ها از طریق لایه شبکه وارد می-شوند که به آن لایه ورودی گفته می‌شود و سپس به مسیر خود از طریق یک یا چند لایه پنهان ادامه می‌دهند و در نهایت از طریق لایه خروجی خارج می‌شوند. گره‌های موجود در هر لایه به گره‌های موجود در لایه های بعدی متصل هستند. هر گره در لایه مخفی و لایه خروجی داده‌های مربوط به گره‌های قبلی را تبدیل و به آن یک تابع فعال سازی اعمال می‌کند. مقداری که تابع فعالسازی تولید می‌کند ۰ یا ۱ است (۱ گره را روشن و ۰ گره را خاموش می‌کند) و این مقدار را به گره‌های متصل در لایه‌های بعدی منتقل می‌کند. اتصالات میان گره‌ها وزن دار می‌شوند از این رو ممکن است گره‌ها دچار خطای یکطرفه (بایاس) شوند. برای مثال، اگر سه گره x_1، x_2 و x_3 گره چهار x_4 را با خطای یکطرفه (θ) (بایاس) از طریق نقاط اتصالی که با ضرایب w_1، w_2و w_3وزندار شده اند تغذیه کنند، آنگاه مقداری که x به لایه بعدی منتقل می‌کند برابر $x = f(\sum ix_i w_i + \theta)$ خواهد بود. جزئیات این فرآیند در شکل ۴ نشان داده شده است.

مقادیر وزن‌ها و بایاس‌ها از طریق یک فرآیند آموزشی تعیین می‌شوند. داده‌ها از طریق شبکه وارد می‌شوند و وزن اتصالات و بایاس گره‌ها از طریق یک الگوریتم انتشار بازگشتی تعدیل می‌شوند. فرآیند تغذیه مستقیم و انتشار بازگشتی آنقدر تکرار می‌شود تا خروجی به دقت مطلوب برسد و یا تعداد مشخصی تکرار انجام شده باشد.

در طبقه بندی، اندازه لایه ورودی معمولا تعداد صفات (*attributes*) مجموعه داده‌ها است و اندازه خروجی برابر با تعداد طبقات (دسته‌ها) است. هنگامی که تنها دو طبقه موجود باشد، می‌توان تنها از یک لایه خروجی با مقادیر هدف مشخص استفاده کرد که نشان دهنده طبقه نمونه است. اندازه و تعداد لایه‌های مخفی را می‌توان تغییر داد که خود یکی از معایب BPN محسوب می‌شود. اگرچه دستورالعمل هایی برای تعیین تعداد گره ها و لایه ها وجود دارد، ولی ساختار آن تا حدود زیادی از طریق آزمون و خطا تعیین می‌شود. مطالعه فعلی همزمان از این دو روش (SOM+BPN) استفاده می‌کند تا بهترین نتایج برای پیش‌بینی حاصل شوند.

www.SID.ir

Archive of SID

۳. نتایج تحلیل

برای انجام این تحقیق ۶۰۵ پرسشنامه میان کارمندان تایوانی با سنین مختلف از اول ژانویه ۲۰۰۹ تا ۱۴ آپریل ۲۰۰۹ توزیع کردیم. ۵۱۱ پرسشنامه به دست ما رسید که معادل ۸۵٪ از پرسشنامه‌های توزیع شده می‌شد. بعد از حذف پرسشنامه های فاقد پاسخ و دارای پاسخ‌های بی دلیل ۴۲۱ پرسشنامه معتبر باقی ماند. از این رو، نرخ بازگشت معتبر برابر ۷۰٪ می‌شد. این بخش به بررسی اعتبار نمونه‌های تحقیقاتی، آزمون قابلیت اطمینان و متغیرها می‌پردازد. افزون بر این، در این قسمت نتایج حاصل از خوشه بندی از طریق SOM=BPN، مدل پیش بینی کننده و ارزیابی عملکرد خوشه بندی نیز آمده است.

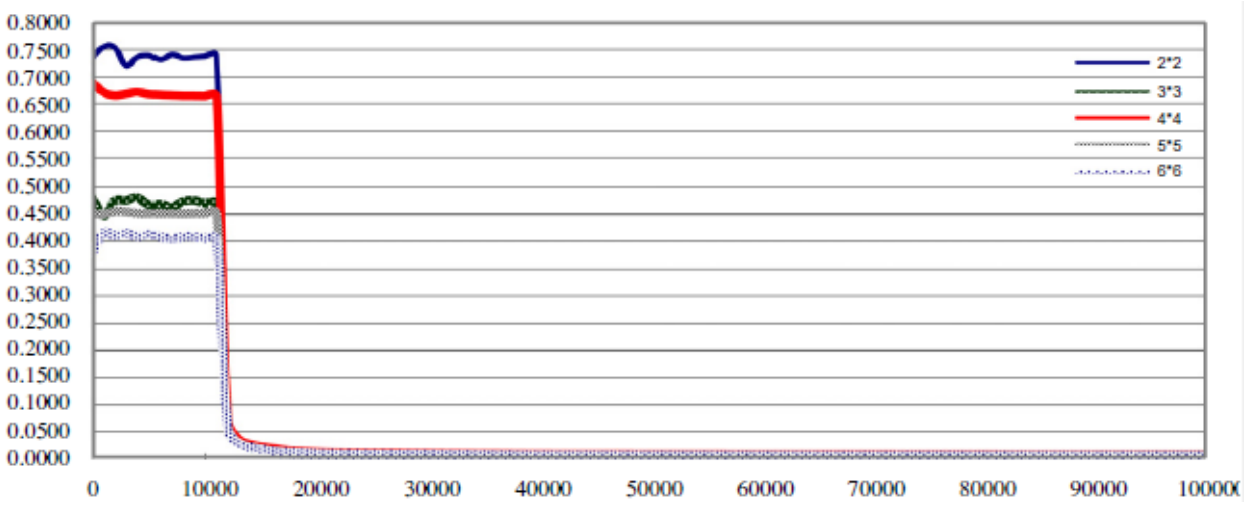
جدول ۳ مقادیر پارامترها در SOM

پارامتر	مقدار
سطح ورودی	۲۸
سطح خروجی	۱
سطح مخفی	۱
قاعده یادگیری	قاعده دلتا
قاعده تبدیل	شبه سیگما

جدول ۴ تحلیل سرعت همگرایی خوشه بندی SOM

گروه خوشه بندی	تعداد دفعات یادگیری	مقدار RMSE (خطای موثر)	سرعت همگرایی
C=2	۷۹۰۰۰	۰,۰۰۹۰۰	
C=3	۷۹۰۰۰	۰,۰۰۸۰۰	
C=4	۵۲۰۰۰	۰,۰۰۸۵۰	بهترین حالت ممکن
C=5	۸۰۰۰۰	۰,۰۰۷۳۰	
C=6	۷۹۰۰۰	۰,۰۰۷۲۰	

www.SID.ir



شکل ۵ منحنی یادگیری در SOM با ۲ تا ۶ خوشه

۴,۱ نتایج حاصل از خوشه بندی SOM

از روش خوشه بندی SOM برای طبقه بندی اولیه گروه‌های خوشه بندی استفاده شده است. در مطالعه فعلی، ۳۶ نمونه به صورت تصادفی از ۴۲۱ نمونه انتخاب کردیم. این ۳۶ نمونه برای آزمایش بکار رفتند. از ۳۸۵ نمونه باقی مانده در مدل SOM برای تحلیل خوشه‌ای استفاده شد. بعد از شناسایی بهترین خوشه ها (که در جدول ۳ و ۴ و شکل پنج آمده است)، روش خوشه بندی k میانگین (k-means) را به تمام مجموعه داده‌ها اعمال کردیم. مطابق پیشنهادات بیان شده در بخش ۳، این تحقیق بر ۲۸ متغیر متمرکز بود (جدول ۲). برای افزایش بازدهی، تصمیم گرفتیم از مقدار RMSE برای شاخص اصلی خوشه بندی استفاده کنیم و مشخص شد که در صورت قرار داشتن دامنه خوشه بندی بین ۲ تا ۶ بهترین نتیجه حاصل می‌شود.

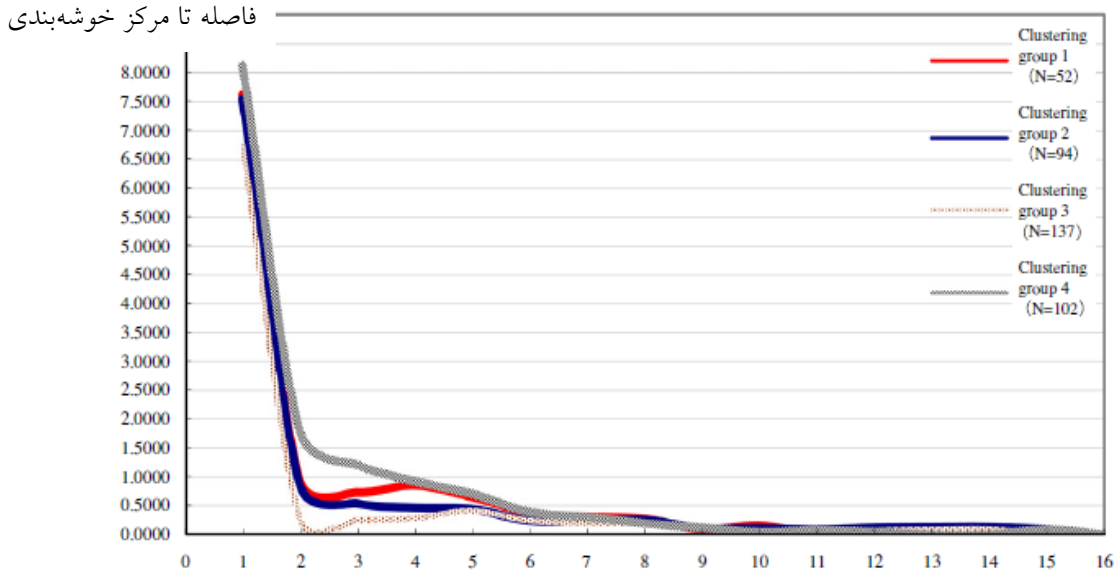
جدول ۲ متغیرهای ورودی

شماره متغیر ورودی	نام متغیر ورودی	شماره متغیر ورودی	نام متغیر ورودی
۱	رشد و خود شکوفایی	۱۵	دسترسی به امور تجاری
۲	رابطه میان یکدیگر	۱۶	داشتن ابتکار عمل در همکاری
۳	حقوق و مزایا	۱۷	فداکاری و وابستگی
۴	محیط کاری	۱۸	داشتن ابتکار عمل برای پذیرش مسئولیت
۵	رهبری و مدیریت	۱۹	تخصیص داخلی سازی
۶	ارتباطات و عدات	۲۰	صیانت از رفاه عمومی
۷	سازگاری با همکاران	۲۱	پیروی
۸	استرس	۲۲	تعهدات و مسئولیت‌ها
۹	خستگی	۲۳	نگرش‌های همکاران و ناظر
۱۰	ناراحتی	۲۴	ساکت ماندن و منتظر مزایا بودن
۱۱	خودباوری پایین	۲۵	سن
۱۲	تداخل با امور خانوادگی	۲۶	تعداد سالهای خدمت

www.SID.ir

۱۳	اذعان به داخلی سازی	۲۷	مسئولیت‌های خانوادگی
۱۴	پیروی محض	۲۸	وضعیت تاهل

شکل ۶ منحنی یادگیری تحلیل خوشه بندی SOM+BPN در چهار گروه



۴،۲ نتایج حاصل از خوشه بندی SOM+BPN

بعد از استخراج گروه خوشه بندی اولیه از SOM، از ANN و BPN برای خوشه بندی تمام داده‌ها استفاده کردیم. همان گونه که در شکل ۶ نشان داده است، همگرایی در ۱۶ تکرار کامل شد که خود منجر به حصول بهترین نتایج شد.

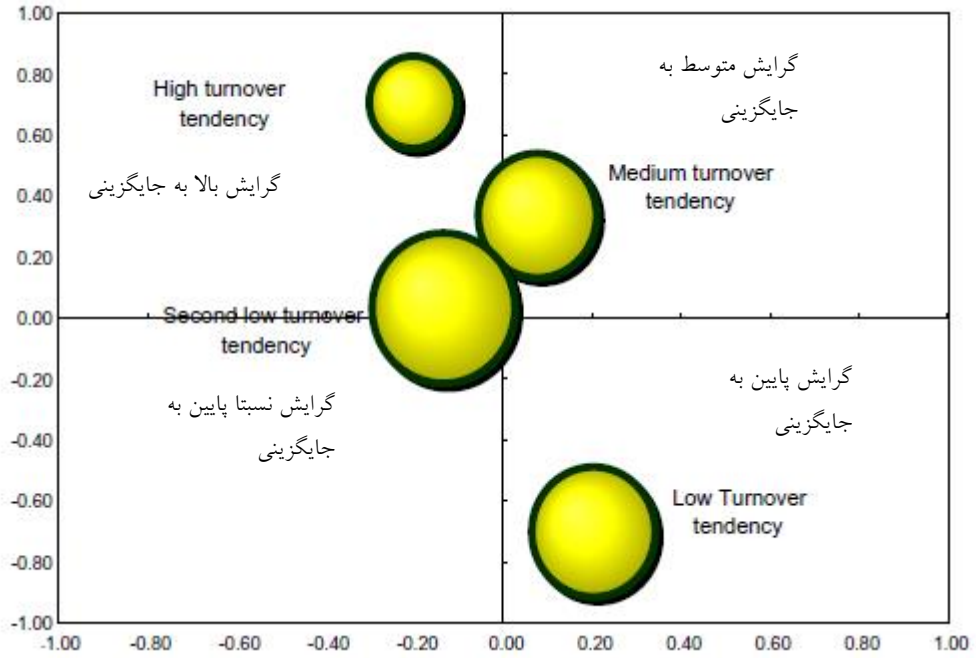
این رویکرد از طریق خوشه بندی دو مرحله‌ایSOM+BPN تمام داده‌ها را با موفقیت در چهار خوشه (گروه) قرار داد: خوشه گرایش بالا به جایگزینی، خوشه گرایش متوسط به جایگزینی، خوشه گرایش نسبتا پایین به جایگزینی و خوشه پایین ترین گرایش به جایگزینی. توزیع این چهار خوشه در شکل ۷ نشان داده شده است. بعد از تکمیل تفکیک این چهار گروه، چهار نمونه به صورت تصادفی برای آزمایش دقت این مدل هیبرید (ترکیبی) انتخاب کردیم.

جدول ۵ نتایج حاصل از پیش بینی داده‌های مورد آزمایش با استفاده از مدل هیبرید را نشان می‌دهد. ستون‌های خاکستری به معنای وجود اشباه در روش طبقه بندی هستند. دقت این مدل بیش از ۹۲٪ بود. افزون بر این، مدل هیبرید (ترکیبی) بهتر از مدل‌های قدیمی از جمله مدل میانگین k و مدل ANN بود. جزئیات نتایج در جدول ۶ آمده است.

www.SID.ir

Archive of SID

شکل ۷ نتایج حاصل از خوشه بندی دو مرحله‌ای (گرایش به جایگزینی)



جدول ۵ سی و شش داده آزمایشی برای پیش بینی نتایج

داده مورد ازمایش	Clustering groups				ارزیابی	
	1	2	3	4		
1	−4.003	−4.453	−2.632	−4.195	3	3
2	−6.358	−0.533	−4.379	−4.495	1	2
3	−10.131	−1.276	−3.820	−2.637	2	2
4	−10.574	−1.672	1.559	−8.895	3	3
5	−7.289	−7.429	−1.756	−1.005	4	4
6	11.278	−3.237	−5.120	−11.033	1	1
7	7.598	−6.674	−3.898	−6.959	1	1
8	−9.820	−0.662	−2.994	−4.473	2	2
9	−11.547	1.861	−0.748	−8.446	1	2
10	−11.430	−6.079	−2.277	0.444	4	4
11	−11.553	−4.284	−2.651	−0.557	4	4
12	−7.173	−7.807	1.066	−4.206	3	3
13	−12.245	−3.574	−1.152	−2.676	3	3
14	−8.135	−4.933	−1.165	−3.445	3	3
15	−1.225	−5.550	−3.756	−3.290	1	1
16	3.897	−5.824	−4.361	−4.855	1	1
17	−4.069	−1.649	−5.494	−3.266	2	2
18	−0.988	0.550	−6.710	−5.541	2	2
19	−9.142	1.260	−5.720	−2.881	2	2
20	−10.428	−2.277	−3.727	−1.679	4	4
21	−7.219	−4.138	−4.358	−0.413	4	4
22	−7.848	−3.252	−0.891	−5.433	3	3
23	−5.830	−2.348	0.750	−9.130	3	3
24	0.302	−4.912	−2.402	−6.115	1	1
25	−0.474	−5.719	−5.349	−2.049	1	1
26	−4.559	−0.268	−2.972	−7.558	2	2
27	−5.418	−0.165	−5.215	−4.348	2	2
28	−11.298	0.020	−9.489	4.074	4	4
29	−12.260	−1.690	−7.733	3.442	3	4
30	−7.532	0.989	−7.978	−0.657	2	2
31	−5.957	−11.819	1.595	−1.639	3	3
32	−9.301	−4.400	−4.758	0.721	4	4
33	−11.886	−1.960	−6.489	2.192	4	4
34	−7.184	−0.889	−5.679	−1.977	2	2
35	−5.249	−1.170	−8.365	0.426	4	4
36	−5.050	−2.734	−6.725	−0.473	4	4

www.SID.ir

Archive of SID

جدول ۶ مقایسه کلی میان BPN، میانگین k و SOM+BPN

روش	میانگین k	BPN	SOM+BPN
دقت پیش بینی نتایج	۶۳,۵	۸۷,۲	۹۲

۴. نتیجه گیری

مطالعات قبلی در حوزه جایگزینی کارکنان بیشتر با هدف بررسی دلایل جایگزینی انجام می‌شدند و به ندرت گرایش به آن را پیش‌بینی می‌کردند. از این رو،استفاده از آنها برای اجتناب از ترک شرکت توسط کارکنان کلیدی به شکلی فوری و موثر امکانپذیر نبود. این مطالعه با استفاده از تحلیل خوشه بندی در داده کاوی این مشکل را حل کرد. با توجه به ویژگی‌های هر خوشه دریافتیم که کارکنان به عوامل مشخصی چون شناسایی داخلی ناظران، پشتیبانی تجاری، رهبری، مدیریت، رشد و خود شکوفایی اهمیت می‌دهند. از لحاظ گرایش به جایگزینی، تاثیر تعهد (یا وفاداری) ناظر بیش از تعهد سازمانی بود که خود یکی از وجوه تمایز این مطالعه با سایر نتایج حاصل از تحقیقات انجام گرفته برروی کشورهای غربی محسوب می‌شود. مدیران نباید تصور کنند که احتمال ترک شرکت توسط کارمندان در صورت تکمیل وظایف آنها و یا افزایش حقوق و مزایا کمتر می‌شود. افزون بر این، هنگامی که شرکت‌ها قوانین و مقررات سختگیرانه ای وضع می‌کنند و به بعد انسانی محیط کار توجه نمی‌کنند، آنگاه ناظران می‌توانند کارمندان رده پایین را دور خود جمع کنند و آنها را به ترک شرکت فعلی و پیوستن به شرکت‌های دیگر ترغیب کنند. مشخصا، تربیت کارکنان وفادار بیش از پیروی از تک تک قوانین اهمیت دارد.

به خاطر ملاحظات زمانی و هزینه‌ای، این مطالعه را می‌توان از لحاظ دقت پیش بینی و کاربرد تحقیقاتی با استفاده از روش‌های دیگر بهبود بخشید. از این رو پیشنهاد می‌کنیم که سایر محققان این حوزه مطالعاتی را انجام دهند که نرخ جایگزینی را براساس قیاس لحاظ نمی‌کند. با این اوصاف، می‌توان از نظریه فازی و تصمیم گیری چند معیاری برای پیدا کردن سایر عوامل موثر در جایگزینی استفاده کرد. به منظور تفکیک بهتر پرسشنامه‌های معتبر، می‌توان از تله‌های نظرسنجی و ارزیابی پرسشنامه‌های معتبر از طریق یک برنامه طراحی استفاده کرد. با این کار صرفه جویی قابل ملاحظه‌ای در وقت می‌شود.

www.SID.ir

منابع

Berardi, V. L., & Zhang, G. P. (1999). The effect of misclassification costs on neural network classifiers. Decision Sciences, 30(3), 659–682.

Dalton, D. R., Todor, W. D., & Krackhardt, D. M. (1982). Turnover overstated: The functional taxonomy. Academy of Management Review, 7, 118–119.

Hao, C. C., Jung, H. C., & Yenhui, O. (2009). A study of the critical factors of the job involvement of financial service personnel after financial tsunami: Take developing market (Taiwan) for example. African Journal of Business Management, 3(12), 798–806.

Hart, S. (1992). Games in extensive and strategic forms. In R. J. Aumann & S. Hart (Eds.), Handbook of game theory with economic applications. Amsterdam: North-Holland.

Kraut, A. I. (1975). Predicting turnover of employees form measured job attitudes. Organizational Behavior and Human Performance, 13, 233–243.

Kuo, R. J., Ho, L. M., & Hu, C. M. (2001). Integration of self-organizing feature map and K-means algorithm for marketing segmentation. Journal of Computers and Operation Research.

Michaels, C. E., & Spector, P. E. (1982). Causes of employee turnover: A test of The Mobley, Griffrth, Hand, and Meglino model. Journal of Applied Psychology, 67(1,(۵۳_۵۹.

Miller, H. E., Miller, H. E., Katerberg & Hulin, C. L. (1979). Evaluation of the Mobley, Horner and Hollingsworth model of employee turnover. Journal of Applied Psychology, 64, 509–517.

Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 62(2) ۲۳۷_۲۴۰.

Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. Psychological Bulletin, ۸۶(۳), ۴۹۳_۵۲۲.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. (1978). An evaluation of Precursors of hospital employee turnover. Journal of Applied Psychology, 63(4)

www.SID.ir

Archive of SID

.٤١٤–٤٠٨

Newman, J. E. (1974). Predicting absenteeism and turnover: A field comparison of Fishbein's mold and traditional job attitude measures. Journal of Applied Psychology, 59, 610–615.

Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. Psychological Bulletin, 80(2), 151–176.

Samuel, M. O., & Chipunza, C. (2009). Employee retention and turnover: Using motivational variables as a panacea. African Journal of Business Management, 3(9), 410–415.

www.SID.ir