ارائه یک مدل تصمیم‌گیری در تعیین تشابه دو وقوع زمین لرزه جهت تسریع در امداد

عارف کریمی افشار[۱]، محمدرضا عظیمی[۲]، معصومه سبکدست[۳]، محمدعلی منتظری[۴]
[۱]دانشگاه صنعتی اصفهان-دانشکده برق و کامپیوتر، A.karimiafshar@ec.iut.ac.ir
[۲] دانشگاه شهید باهنر کرمان -دانشکده برق و کامپیوتر ، M.azimi@eng.uk..ac.ir
[۳]دانشگاه شهید باهنرکرمان-دانشکده برق و کامپیوتر ، M.sabkdast@eng.uk.ac.ir
[۴]دانشگاه صنعتی اصفهان-دانشکده برق و کامپیوتر ، Montazeri@cc.iut.ac.ir

چکیده - تصمیم گیری دارای ماهیتی ادراکی و غیر کمی می‌باشد لذا در تصمیم سازی با استفاده از سیستم های کامپیوتری که ذاتا دارای ماهیتی ساختاریافته و غیرادراکی می باشد، با مشکل مواجه خواهیم بود. تلاش های صورت گرفته در قالب سیستم های پشتیبانی تصمیم با استفاده از تکنیک‌های هوش‌مصنوعی و داده‌کاوی، سعی در بر طرف کردن این مسئله به روش های متعددی داشته است. از سوی دیگر وقوع زمین لرزه امری غیرقابل پیش‌بینی بوده و اجتناب‌ناپذیر می‌باشد. تصمیم‌گیری در شرایط وقوع زمین‌لرزه تحت تاثیر وخامت اوضاع با مشکل روبه‌رو خواهد بود، امداد رسانی به مجروحین، توزیع مناسب نیروها، تشخیص به موقع نقاط بحرانی از جمله اولین اقدماتی است که باید پس از وقوع یک زمین لرزه صورت گیرد. در این مقاله با ارایه یک روش ترکیبی جدید، سعی شده‌است که با توجه به وقوع زمین لرزه های قبلی بهترین تطابق با زمین لرزه جاری تشخیص داده شود تا در کوتاه‌ترین زمان ممکن نقاط بحرانی شناخته شده و امداد رسانی شود. نتایج حاصل از پیاده‌سازی روش پیشنهادی حاکی ازصحت ۸۷/۲ درصدی تصمیمات با تقریب کمتر از ۰/۱۲ بوده‌است.

کلید واژه- تصمیم گیری چند معیاره، سیستم‌های پشتیبانی تصمیم، داده‌کاوی، زمین‌لرزه.

۱- مقدمه

در دنیای واقعی اخذ تصمیم، یک فرایند ساده یک بعدی به شمار نمی‌آید، حتی در ساده‌ترین حالات نیز یک تصمیم‌گیری با پارامترها و معیارهای متعددی رو به‌رو خواهد بود. در جواب به این واقعیت روش‌های حل یا اخذ تصمیمات چند معیاره مطرح گردیده است. یکی از جامع ترین مدل‌های طراحی شده برای تصمیم‌گیری با معیارهای چند گانه، فرآیند تحلیل سلسله مراتبی می باشد[۱]، زیرا این تکنیک امکان فرموله کردن مسئله را به صورت سلسله مراتبی فراهم می‌کند و همچنین امکان در نظر گرفتن معیارهای مختلف کمی و کیفی را در مسئله دارد. این فرایند گزینه‌های مختلف را در تصمیم گیری دخالت داده و امکان تحلیل حساسیت معیارها و زیر معیارها را دارد. علاوه بر این، بر مبنای مقایسه زوجی بنا نهاده شده، که قضاوت و محاسبات را تسهیل می‌نماید. همچنین میزان سازگاری و ناسازگاری تصمیم را نشان می‌دهد که از مزایای ممتاز این تکنیک در تصمیم‌گیری چند معیاره می‌باشد. بعلاوه از یک مبنای تئوریک قوی برخوردار بوده و بر اساس اصول بدیهی بنا شده است، در این فرایند تصمیم گیری از مقایسه زوجی جهت شکل‌دهی ماتریس مبدل داده‌های کیفی به مقادیر عددی استفاده می شود.

روش ایده‌آل کلاسیک، در این روش با توجه به معیارها و گزینه های موجود، ماتریس تصمیم ساخته می شود و پس از ارزش گذاری معیارها و ایجاد ماتریس وزن‌ها، ماتریس تصمیم وزن‌دار ساخته شده و با حل گزینه ایده آل کلاسیک گزینه مناسب بر حسب فاصله از ایده‌آل یا ضد ایده‌آل انتخاب می شود[۲]. روش دلفی، در این روش از دانش افراد خبره جهت رسیدن به مطمئن‌ترین توافق گروهی، زمانی که دانشی نامطمئن و ناکامل در یک موضوع خاص در دسترس باشد، استفاده می شود، روند کار از طریق استفاده از پرسشنامه و نظر خواهی از متخصصین امر و افراد خبره در طی دفعات و باز خورد حاصل از آن، دنبال می شود[۳].

اما اصولا در بسیاری از تصمیم‌گیری‌ها زمان یک پارامتر اصلی به شمار نمی‌آید، برای مثال در فرایند تصمیم‌گیری ساخت

یک سد خاکی صرف زمان یک ماه چندان قابل اهمیت نمی‌باشد، این در حالی است که تصمیم‌گیری در شرایط بحران، برای مثال رخداد یک زمین‌لرزه، هر دقیقه در فرایند تصمیم‌گیری دارای اهمیت می‌باشد. بنابراین عملا حوزه‌ی روش‌های قابل استفاده محدود خواهد شد، برای مثال استفاده از روش دلفی[۳] در یک تصمیم‌گیری شرایط وقوع زلزله نتنها مفید نخواهد بود حتی ممکن است جان انسان‌های زیادی را به خطر بیندازد، هر چند که ممکن است تصمیم نهایی درصد بالایی از اطمینان یا صحت را به همراه داشته باشد. بنابراین شرایط بحران روش‌هایی را طلب می‌کند که کاملا مبتنی بر سیستم‌های کامپیوتری بوده، بر خلاف روش‌های مانند دلفی، تا در سریع‌ترین زمان ممکن تصمیمات لازم اتخاذ و ابلاغ گردد. همچنین این روش‌ها باید انعطاف پذیر بوده و از قابلیت اطمینان بالایی برخوردار باشند.

در این مقاله سعی شده است با ایجاد تغییر در روش Apriori از سابقه رخ دادهای قبلی زلزله بهره‌گیری شود و با تولید مجموعه قوانین مرتبط در مورد نقاط بحرانی، تعداد افراد تحت تاثیر قرار گرفته، یا میزان خسارات در راستای امدادرسانی تصمیم گیری شود. در ادامه ساختار این مقاله به ترتیب زیر خواهد بود، در بخش دوم روش Apriori که مبنای تصمیم گیری طرح پیشنهادی می باشد بیان می گردد، بخش سوم توضیح مختصری از روش خوشه‌بندی پیشنهادی را در بر خواهد داشت، بخش چهارم روش پیشنهادی در این مقاله را بیان می‌نماید، بخش پنجم به شبیه سازی اختصاص دارد و در بخش ششم نتیجه‌گیری مقاله آورده شده است.

۲- روش Apriori

روش Apriori ، روش متداول کاوش در یک مجموعه با آیتم‌های متناوب بوده که بر مبنای بیش‌ترین تکرار یک آیتم یا مجموعه‌های پرتکرار چند عضوی از زیر مجموعه‌های مجموعه‌ی اصلی، قوانین مورد نیاز کاوش را تولید می‌نماید. الگوریتم Apriori به روش زیر عمل می‌کند.

۱- پایگاه تراکنش را یک بار پویش کن تا L_1، یعنی مجموعه همه الگوهای یک آیتمی، را پیدا کن.

۲- به ازای $(k=2; L_{k-1} \neq \Phi; k++)$ عملیات زیر را تکرار کن:

a) مجموعه کاندیداهای k آیتمی یعنی C_k را تولید کن. مجموعه آیتم k-آیتمی X در C_k خواهد بود. اگر و تنها اگر هر زیرمجموعه k-1 عضوی از X در L_{k-1} باشد.

b) اگر $C_k \neq \Phi$ است به مرحله ۳ برو.

c) برای شمارش تعداد هر یک از مجموعه آیتم های موجود در C_k، پایگاه تراکنش را یک بار پویش کن.

d) L_k را از طریق رابطه (۱) بروزرسانی کن.

$$L_k = \{X \mid (X \in C_k) \wedge (SUP(X) \geq \min\text{-}\sup)\} \qquad (۱)$$

۳- مقدار $\cup_{i=1}^{k} L_i$ را بازگردان.

الگوریتم Apriori نخستین الگوریتم غیربدیهی است که برای کاوش مجموعه آیتم‌های متناوب ساخته شده است. پس از آن الگوریتم‌ها و روش‌های زیادی برای بهبود کارآیی این الگوریتم ارائه گردید. اما این روش دارای دو مشکل ذاتی است که مستقل از نحوه پیاده سازی آن بوده و بنابراین در همه بهبودهای صورت گرفته بر روی آن نیز این مشکل برقرار و باقی است:

- تعداد کاندیداهایی که برای یافتن مجموعه آیتم‌های متناوب در این روش ساخته می‌شود، می‌تواند فوق‌العاده زیاد باشد مخصوصا اگر طول مجموعه آیتم‌های متناوب بزرگ باشد. به عنوان مثال برای کاوش الگوی متناوبی با طول ۱۰۰ باید $\binom{100}{1}$ کاندیدای یک آیتمی، $\binom{100}{2}$ کاندیدای دو آیتمی و... $\binom{100}{100}$ کاندیدای صد آیتمی ساخته شود. بنابراین در مجموع فوق روش $\binom{100}{1}+\binom{100}{2}+\ldots+\binom{100}{100}=2^{100}-1>10^{30}$ کاندیدا تولید می‌نماید. هزینه تولید مزبور، هزینه ذاتی تولید کاندیداها است و به الگوریتمی که آن را پیاده‌سازی می‌کند ربطی ندارد.
- پویش مکرر پایگاه تراکنش‌ها و چک کردن مجموعه بسیار بزرگی از کاندیداها برای تست انطباق الگوها، بسیار زمان‌گیر است و این موضوع به خصوص در مورد کاوش الگوهای بزرگ صادق است.

بنابراین تعداد زیاد کاندیداها و پویش چندین باره مجموعه داده، سبب ایجاد محدودیت در کارآیی الگوریتم‌های مبتنی بر جستجوی اول سطح می‌گردد. ما برای رفع این مشکل از تکنیک خوشه‌بندی K-means استفاده نمودیم تا ضمن کاهش حجم داده‌ها امکان بهره‌گیری از داده های عددی مربوط به زمین لرزه-های قبلی نیز استفاده نماییم.

۳- خوشه‌بندی به روش K-means

این روش علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش روشی انحصاری و مسطح محسوب می‌شود.[۱۰] برای این الگوریتم شکل‌های مختلفی بیان شده است. ولی همهٔ آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

- بدست آوردن نقاطی به عنوان مراکز خوشه‌ها، این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.
- نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد.

در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز نقاطی به صورت تصادفی انتخاب می‌شود. سپس در داده‌ها با توجه با میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آن‌ها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود. تابع رابطه (۲) به عنوان تابع هدف مطرح است.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| X_i^{(j)} - c_j \right\|^2 \qquad (۲)$$

که $\| \|$ معیار فاصلهٔ بین نقاط و c_j مرکز خوشهٔ j ام است.
الگوریتم زیر الگوریتم پایه برای این روش محسوب می‌شود:

۱. در ابتدا K نقطه به عنوان نقاط مراکز خوشه‌ها انتخاب می‌شوند.
۲. هر نمونه داده به خوشه‌ای که مرکز آن خوشه کمترین فاصله تا آن داده را داراست، نسبت داده می‌شود.
۳. سپس تعلق تمام داده‌ها به یکی از خوشه‌ها برای هر خوشه یک نقطه جدید به عنوان مرکز محاسبه می‌شود. (میانگین نقاط متعلق به هر خوشه)
۴. مراحل ۲ و ۳ تکرار می‌شوند تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود.

۴- روش پیشنهادی

روش Apriori اصولا جهت کاوش در یک مجموعه مورد استفاده قرار می گیرد. اگر در سطح بالاتری نگاه کنیم همچنین زمانی مورد استفاده قرار می‌گیرد که بخواهیم از وقوع یا عدم وقوع یکسری حوادث یا رخدادها، تخمینی از رخدادهای دیگر تولید کنیم. در این جا ما مسئله تصمیم گیری در شرایط وقوع زمین لرزه جهت تعیین نقاط بحرانی یا تعداد افراد تحت تاثیر قرار گرفته را به یک مسئله تخمین بر اساس شبیه‌ترین گزینه تبدیل می‌کنیم. فرض کنید در K رخداد زمین لرزه قابل مقایسه مجموعه قوانین A تولید شده است. حال در تعیین نقاط بحرانی می توان از این قواعد استفاده نمود. برای مثال اگر زمین لرزه‌ای به قدرت ۴/۸ ریشتر و در عمق ۳۵ متری رخ داده باشد و بخواهیم افراد تحت تاثیر را مشخص کنیم. به همه قواعد تولید شده در مجموعه قواعد حاصل از مجموعه داده با Support بالای ۷۰ درصد که قدرت بر حسب ریشتر در یک سمت و افراد تحت تاثیر در طرف دیگر قرار دارند یا قوانینی با Confidnce بالای ۸۵ درصد که قدرت بر حسب ریشتر در سمت چپ قرار دارد توجه می‌کنیم. برای مثال یک قاعده بیان می‌کند که زمین لرزه هایی با میزان بزرگی بر حسب ریشتر ۴/۸ و عمق ۳۵ متری تعداد افراد تحت تاثیر بین ۱۰۰ تا ۴۵۰ نفر خواهد داشت. روش پیشنهادی به این صورت عمل خواهد کرد که ابتدا داده‌های موجود در رکوردهای قابل مقایسه را جدا نموده و به تعداد خوشه های از قبل مشخص شده‌ای تبدیل می‌نماید. سپس این خوشه-ها را به عنوان پارامترهای اصلی در نظر گرفته که مقدار آن‌ها می تواند درست یا غلط باشد. سپس همه خوشه‌های پارامترهای مختلف را در نظر گرفته و به الگوریتم Apriori می دهیم. اعمال Apriori بر این مجموعه داده قوانینی تولید خواهد کرد که بر اساس آن‌ها می توانیم پارامترهای دیگر موردنظر را تخمین بزنیم.

۵- شبیه سازی

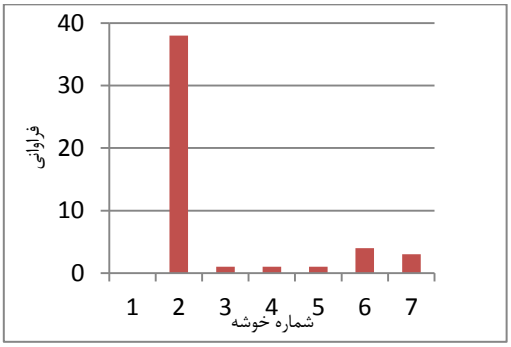
پس از تنظیم روش Apriori در نرم افزار کلمانتین[1]، از مجموعه داده disaster-data-collection در بانک های اطلاعاتی Emdat [۹]جهت شبیه‌سازی و تست استفاده شد[۸]. disaster-data-collection یک مجوعه داده در مورد وقوع بلایای طبیعی طی دوره‌های مختلف زمانی در کشورهای مختلف از جمله ایران، در حوزه های مختلفی همچون خشکسالی، بیماری‌های واگیر، زمین‌لرزه و ... می باشد. که توسط (disaster-data-collection-initiative-portal) آنالیز و ارائه گردیده است. این مجموعه داده در حوزه‌ی زمین‌لرزه حاوی اطلاعاتی در مورد تعداد کشته‌ها، میزان خسارات، تاریخ دقیق وقوع، محل و همچنین بزرگی و سایر ویژگی‌های آن زمین لرزه می‌باشد. مسئله تصمیم‌گیری به صورت ارائه یک مجموعه قواعد جهت تعیین نقاط بحرانی، افراد تحت تاثیر قرار گرفته و میزان خسارت جهت تسهیل در امداد، تعریف شده‌است. دسته‌بندی خوشه‌ها، بر مبنای ترکیب‌های مختلف تعداد کشته‌ها، میزان خسارت و قدرت زمین‌لرزه و اطلاعات عمق زمین لرزه صورت گرفته است.

با توجه به حجم نسبتا زیاد داده‌ها و ناکارآمدی روش-های کلاسیک، برای مقایسه روش پیشنهادی از روش دلفی استفاده شد، که امروزه نیز عموما از صورت‌های آماتور همین روش به صورت ضمنی استفاده می‌گردد. جهت تعیین نتایج روش دلفی از دو گروه متخصص نظرخواهی شد.

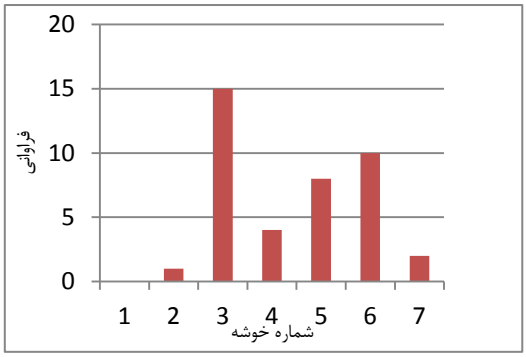
همچنین ارزیابی‌های صورت گرفته برای هر گروه باتوجه به معیارهای مختلف به‌گونه‌ای صورت داه شده است که تمامی معیارها به صورت مستقل در نظر گرفته شوند، هر چند که عدم رعایت این نکته در جواب نهایی چندان تاثیری نخواهد داشت.

نمودار شکل(۱) بیان کننده نحوه توزیع رخدادهای مختلف زمین لرزه در شش خوشه از قبل تعیین شده براساس بزرگی زمین لرزه، نمودار شکل(۲) بیان کننده نحوه توزیع رخدادهای مختلف زمین لرزه در شش خوشه از قبل تعیین شده براساس تعداد کشته‌های زمین لرزه، می باشد، همچنین نمودار های شکل‌های (۳) و (۴) به ترتیب بیان کننده نحوه توزیع رخدادهای مختلف زمین لرزه در شش خوشه از قبل تعیین شده براساس میزان خسارات و عمق کانون زمین لرزه می‌باشد.

شکل ۱: توزیع خوشه‌ها بر حسب بزرگی زمین‌لرزه

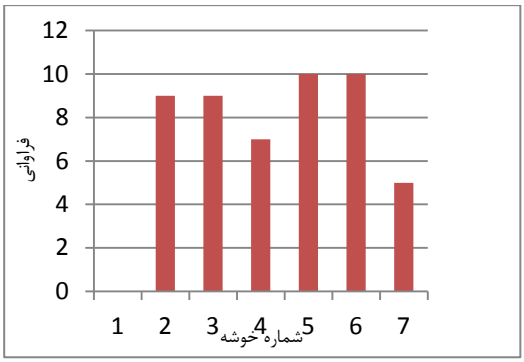


شکل ۲: توزیع خوشه‌ها بر حسب تعداد کشته‌ها زمین‌لرزه



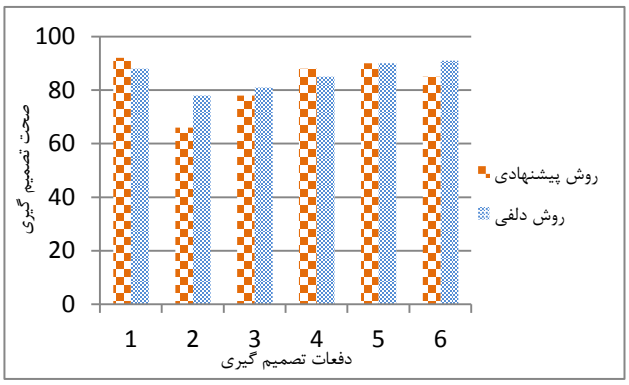
شکل ۳: توزیع خوشه‌ها بر حسب میزان خسارات زمین‌لرزه

[1] - Clementine



شکل ۴: توزیع خوشه‌ها بر حسب عمق کانون زمین‌لرزه

نمودار شکل(۵) مطابقت تصمیمات اتخاذ شده توسط روش پیشنهادی را در مقایسه با روش دلفی برای مجموعه داده مورد استفاده، بیان می‌کند.



شکل ۵ : میزان مطابقت دنباله خروجی در روش های مختلف با دنباله قطعی

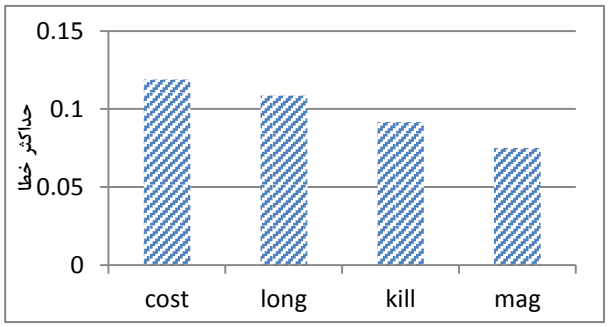
همانطوری که در نمودار شکل (۵) مشخص شده است روش ارائه شده در این مقاله، دارای مطابقت ۸۷/۲ درصدی با دنباله قطعی، می‌باشد. این درحالی است که تولید این دنباله از تصمیمات در کمترین زمان ممکن و بدون نیاز به حضور متخصصین یا نهادهای مسئول صورت می‌گیرد در حالی که روش دلفی مبتنی بر نظرات افراد مسئول یا متخصص و لزوما حضور آن‌ها می‌باشد.

نمودار شکل (۶) بیانگر حداکثر خطا در تصمیمات برحسب پارامترهای مختلف، میزان خسارات، اطلاعات فاصله، تعداد کشته‌ها و بزرگی زمین‌لرزه را نمایش می‌دهد. همان‌طور که مشاهده می‌شود حداکثر خطا از ۰/۱۲ کمتر می باشد.

۶- نتیجه‌گیری

در این مقاله بر مبنای روش‌های تصمیم گیری چند معیاره، روش ترکیبی جدیدی با استفاده از روش Apriori و خوشه‌بندی K-means مطرح گردید، در موضع سنتی زمان یک پارامتر اصلی در تصمیم گیری به شمار نمی‌آمده است و این استفاده از روش‌های کلاسیک را در شرایط وقوع زمین‌لرزه ناکارآمد می‌سازد. با استفاده از روش پیشنهادی تصمیمات با توجه به رخدادهای قبلی زمین‌لرزه در همان محل یا محل‌های مشابه در کمترین زمان ممکن صورت می‌گیرد.

روش پیشنهادی با خوشه‌بندی داده‌های قبلی و مقداردهی هر دسته مسئله تصمیم گیری را به یافتن بالاترین تطابق بین رخدادهای مختلف زمین‌لرزه تبدیل می‌نماید. خوشه‌های تعیین شده توسط روش K-means مورد استفاده روش Apriori قرار می‌گیرد تا منجر به تولید یکسری قوانین جهت تصمیم گیری شود. با استفاده از قواعدی که از این روش حاصل می‌شود، می‌توان نقاط بحرانی، افراد تحت تاثیر را در کم‌ترین زمان ممکن با درصد تقریب مناسب تعیین نمود. پیاده سازی روش پیشنهادی و اعمال آن بر مجموعه داده disaster-data-collection ، حاکی از صحت ۸۷/۲ درصدی در تصمیمات با تقریب کمتر از ۰/۱۲ است.



شکل ۶: میزان حداکثر خطا بر حسب معیار های مختلف

سپاسگزاری

با سپاس فراوان از شرکت ایده‌کارکویرکارمانیا مستقر در مرکز رشد پارک فناوری استان کرمان که از این مقاله حمایت نمودند تا به صورت کاربردی در نرم‌افزار این شرکت مورد استفاده قرار گیرد.

مراجع

[1] S.O. Vaidya, and S. Kumar, "Analytic hierarchy process: An overview of applications", *European Journal of Operational Research 169*, PP.1-29, 2006.

[6] Turban, E., Aronson, J.E., Liang, T.P. *Decision Support Systems and Intelligent Systems*. Prentice Hall, 2005.
[7] J.P. Shim and M. Warkentin, “Past, present, and future of decision support technology” *Decision Support Systems*, Vol. 33, pp. 111-126, 2002.
[8] Spss, “SPSS Neural Networks™ 17.0,” *Analysis*, 2007.
[9] Infochimps Research Center. disaster-data-collection Data set, [Online]. Available: http://infochimps.com/datasets/earthquake, 2011.
[10] R., Agrawal, and R., Srikant., "Fast algorithms for mining association rules." In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, vol. 1215, pp. 487-499. 1994.
[2] G.S. Liang, and M.J. Wang, "Fuzzy MCDM based on ideal and anti-ideal concepts" *European Journal of Operational Research, Vol. 112*, pp. 682-691, 1999.
[3] T., Gal and T. ,Hanne, “Nonessential objectives within network approaches for MCDM” *European Journal of Operational Research*, Vol. 168, pp. 584-592, 2006.
[4] K. Genova, V. Vassilev, F. Andonov, M. Vassileva, and S. Konstantinova," A Multicriteria Analysis Decision Support System", *International Conference on Computer Systems and Technologies*, IIIA 1-6, 2004.
[5] Andriole, S., *Handbook of Decision Support Systems*, TAB Books, Inc., 1989.