پژوهشکده بیمه

رتبه بندی ویژگی‌های اسمی بیمه‌نامه‌های رشته بیمه بدنه اتومبیل با استفاده از اعداد فازی

مهناز منطقی پور،
مسعود منطقی پور، سحر کمالخانی

آبان ۱۳۹۸

بسم الله الرحمن الرحيم

موضوع

به منظور پیاده سازی مدل‌های یادگیری ماشین تعریف فاصله بین اشیاء و نحوه تعریف آن بسیار مهم است. مقیاس ها توسط استیونس به چهار گروه طبقه بندی می‌شوند:

۱- اسمی،
۲- ترتیبی،
۳- فاصله‌ای و
۴- نسبی.

در واقع این چهار طبقه بر اساس میزان قابل قیاس بودن فواصل تعیین شده است [۱].

توصیف داده ها

بیمه نامه های بدنه اتومبیل طی هشت سال:

سال	تعداد بیمه نامه
۱۳۸۶	۱۳۸۱۷۳
۱۳۸۷	۱۳۶۰۹۷
۱۳۸۸	۱۶۴۳۷۱
۱۳۸۹	۱۷۶۸۶۳
۱۳۹۰	۱۹۵۷۸۷
۱۳۹۱	۲۰۹۷۰۸
۱۳۹۲	۱۹۳۴۰۴
۱۳۹۳	۱۷۸۶۴۹

مساله تنوع ویژگی اسمی

تنوع خصوصیات اسمی: (۲۸ نوع پراید)

نام	فراوانی	نام	فراوانی
پراید ۱۳۱	۱۵۱۰۲	پراید ۱۳۲ (۵نفر)	۱۵۴۸۳
پراید ۱۳۱ اس ال	۱۵۱۴۸	پراید ۱۳۲ ایی ایکس	۱۶۱۳۲
پراید ۱۳۱ اس ال (۵نفر)	۱۵۲۸۹	پراید ۱۴۱	۱۰۴۲۵
پراید ۱۳۱ اس ایکس	۱۵۱۰۷	پراید ۱۴۱ اس ال	۱۵۰۷۲
پراید ۱۳۱ اس ایکس (۴نفر)	۱۵۳۹۱	پراید ۱۴۱ آی	۱۱۵۱۸
پراید ۱۳۱ اس ایی	۱۶۱۰۷	پراید ۱۴۱ آی (۵نفر)	۱۵۹۳۹
پراید ۱۳۲	۱۳۴۲۴	پراید ۱۴۱ دی ال ایکس آی	۱۳۰۶۴

در مجموع تعداد انواع حالت های ویژگی نوع خودرو ۴۵۷ و انواع حالت های شهر محل صدور بیمه نامه ۵۸۷ حالت است.

نرخ خسارت

نرخ خسارت را به صورت مجموع مبالغ خسارت هر بیمه‌نامه به سرمایه بیمه شده، آن بیمه‌نامه محاسبه می‌کنیم این شاخص، اثر تورم را در محاسبات کاهش می‌دهد. از طرفی برای مقایسه رفتار خسارت اتومبیل‌ها با قیمت های متفاوت، نمی توان مبالغ خسارت‌ها را مقایسه کرد، اما این شاخص این امکان را فراهم می‌کند تا بتوان رفتار خسارت اتومبیل‌های مختلف با قیمت‌های متفاوت را مقایسه کرد.

نرخ خسارت یک متغیر تصادفی است.

رتبه بندی نرخ خسارت نیازمند روشی برای رتبه بندی متغیر های تصادفی است.

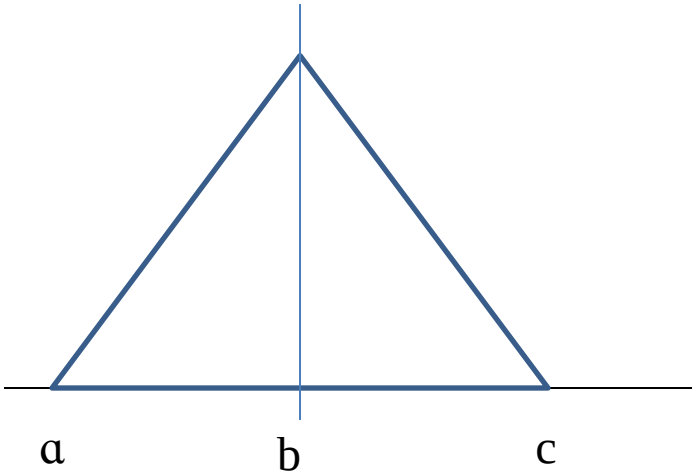
ویژگی های مورد نیاز روش رتبه بندی متغیرهای تصادفی نرخ خسارت

دو معیار را در محاسبات در نظر داشته باشد:
۱- مقادیر متغیر تصادفی
۲- احتمال رخداد مقادیر متغیر تصادفی

روش پیشنهادی:
روش انتخابی، تعریف اعداد فازی با استفاده از نرخ خسارتها و رتبه بندی اعداد فازی

تعریف عدد فازی با استفاده از نرخ خسارت ها

$$\mu(x)=\begin{cases}\dfrac{x-a}{b-a} & a<x<b\\ \dfrac{a-x}{c-b} & b<x<c\\ ۱ & x=b\\ ۰ & o.w\end{cases}$$



پاکسازی داده ها

یکی از چالش‌هایی که در پردازش داده‌های بیمه نامه‌ها وجود دارد، الحاقیه‌های بیمه نامه‌ها هستند. الحاقیه‌ها را نمی‌توان از مجموعه داده‌ها کنار گذاشت و در عین حال نمی‌توان آنها را به عنوان یک ریسک مجزا در نظر گرفت. در این مقاله برای فائق آمدن بر این چالش تاریخ شروع هر الحاقیه به عنوان تاریخ پایان بیمه نامه و یا الحاقیه قبلی در نظر گرفته شده است و هر الحاقیه به منزله یک بیمه نامه در نظر گرفته شده است.

میانگین اعداد فازی

در این میانگین، وزن‌ها نمی‌توانند بر اساس تعداد بیمه نامه‌های هر سال باشند، زیرا تاریخ شروع و پایان بیمه نامه ها یکسان نیست. بنابراین بر اساس تعداد روزهایی که از حالت ویژگی مورد نظر تعهد بیمه‌ای داشته‌ایم، استفاده می‌کنیم. به این ترتیب، میانگین وزنی هشت صدک پنجم و هشت میانگین و هشت صدک نود و پنجم سه مقدار جدید می‌دهند که با آنها عدد فازی جدیدی ساخته-ایم و آن را عدد فازی میانگین می‌نامیم.

$$a^c = \frac{\left(\sum_{j=1}^{9} {w_j}^c \, {p_{0.05}^j}^c\right)}{\sum_{j=1}^{9} {w_j}^c}$$

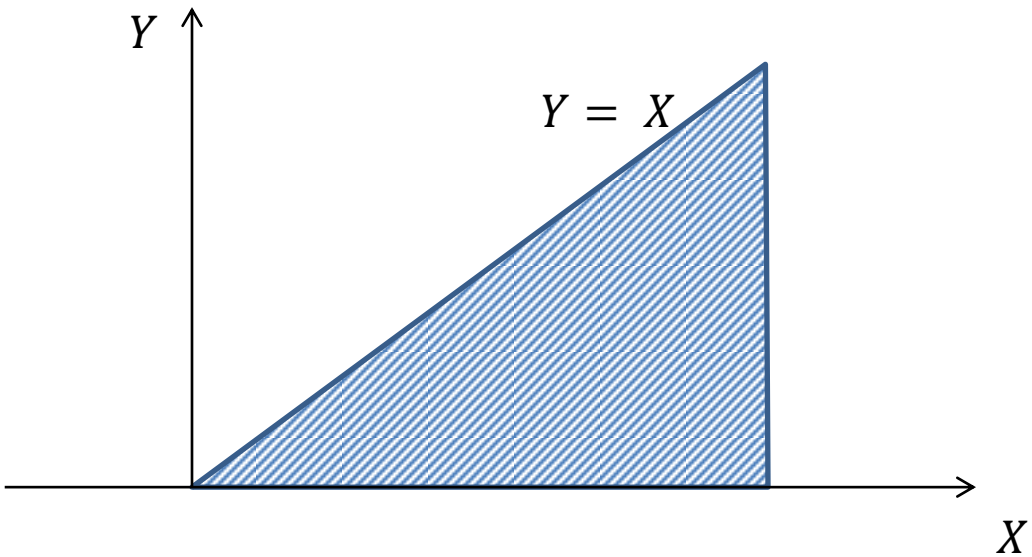
$$m^c = \frac{\left(\sum_{j=1}^{9} w_j \, {m^j}^c\right)}{\sum_{j=1}^{9} {w_j}^c}$$

$$b^c = \frac{\left(\sum_{j=1}^{9} w_j \, {p_{0.95}^j}^c\right)}{\sum_{j=1}^{9} {w_j}^c}$$

روش رتبه بندی اعداد فازی

احتمال بزرگتر بودن متغیر تصادفی X با مقادیری از متغیر تصادفی Y :

$$P(X \geqslant Y) = \int_{-\infty}^{\infty} p_X(x) \left[\int_{-\infty}^{x} p_Y(y) dy \right] dx$$



روش رتبه بندی اعداد فازی

سپس مقدار $P(A \geqslant B)$ که به بزرگتر بودن عدد فازی A از عدد فازی B تفسیر می‌شود، به صورت رابطه زیر تعریف می‌شود:

$$P(A \geqslant B) = \int_0^1 P(A_\alpha \geqslant B_\alpha) d\alpha$$

$$A_\alpha = \{x \epsilon R; \mu_A > \alpha\}$$

فرض کنیم عدد فازی هدف T ، انتخاب شده باشد و می‌خواهیم اعداد فازی $A_1, A_2, \ldots, A_N$ را رتبه بندی کنیم، مقادیر

$$E_T(A_i) = P(A_i \geqslant T) \quad i = 1,2, \ldots, N$$

را محاسبه می کنیم.

روش رتبه بندی اعداد فازی

در پایان شاخص نسبی

$$R_T(A_i) = \frac{E_T(A_i)}{max_{A_j \in S}\{E_T(A_j)\}}$$

برای تعیین رتبه، که عددی بین صفر و یک است تعریف می‌شود.

که برابر با اعداد فازی مثلثی

$$T_{optimistic} = [x_{min}, x_{min}, x_{max}]$$
$$T_{pesimistic} = [x_{min}, x_{max}, x_{max}]$$

محاسبه می‌شوند. که در آن x_{min} و x_{max} برابرند با:

$$,x_{min} = min\{x|\exists A_i\ ;\ \mu_A(x) > 0\}$$
$$.x_{max} = max\{x|\exists A_i\ ;\ \mu_A(x) > 0\}$$

تابع عضویت عدد فازی مستطیلی $T_{neutral}$ در بازه $[x_{min}, x_{max}]$ برابر با ۱ است و خارج از این بازه برابر با صفر است.

نتایج

خودروهای با کمترین رتبه عدد فازی:

نوع خودرو	رتبه
بی ام و ۵۲۸ آی	۰,۰۰۸۱۹۲
بی ام و ۲۸ آی - ایکس ۳	۰,۰۰۸۳۶۶
پورشه کاین	۰,۰۰۹۹۲۴
بنز سی ۲۰۰ سی جی آی	۰,۰۱۰۷۰۳

نتایج

خودرو های با بیشترین رتبه عدد فازی :

نوع خودرو	رتبه عدد فازی
جیپ صحرا کالسکه ای	۰,۴۷۳۵۴۲
سپند پی کی آی	۰,۴۹۵۳۰۷
پیکان ۱۶۰۰	۰,۵۱۷۱۰۵
پیکان دولوکس ۱۶۰۰	۰,۵۲۵۷۶۱
پیکان	۰,۵۲۹۳۱
سپند پی کا	۰,۵۵۳۰۶۸
پیکان دولوکس	۰,۵۹۱۶۳۴
پیکان ۱۸۰۰	۰,۶۰۰۸۶۴
جیپ لندرور	۰,۶۰۴۰۷۴
رنو سپند	۰,۷۱۱۷۷۲
رنو ۵	۱

نتایج

شهرهای با کمترین رتبه عدد فازی :

نام شهر	رتبه عدد فازی
چابکسر	۰,۲۶۲۳۱۸
شهرکرد	۰,۲۶۴۱۹۳
بندرخمیر	۰,۲۷۱۱۰۷
یاسوج	۰,۲۸۸۵۳۸
قره ضیاءالدین	۰,۲۹۸۱۱۳

نتایج

شهرهای با بیشترین رتبه عدد فازی :

نام شهر	رتبه عدد فازی
قیروکارزین	۰,۸۵۹۴۸۹
اقلید	۰,۸۹۶۶۶۵
زرقان	۰,۹۱۸۳۶۸
شادگان	۰,۹۵۶۱۲۱
بجستان	۱

مراجع

[1] J. Han, M. Kamber and J. Pei, Data mining concepts and techniques, Elsevier, 2012 .

[2] N. Dalvi and D. Suci, "Efficient Query Evaluation on Probabilistic Databases," vol. 16, no. 4, pp. 523-544, 2007 .

[3] J. Li and A. Deshpande, "Ranking continuous probabilistic data sets," Proceedings of the VLDB Endowment, vol. 3, no. 1-2, pp. 638-649, 2010 .

[4] J. Li, B. Saha and A. Dashpande, "VLDB Journal manuscript No.(will be inserted by the editor)A Unified Approach to Ranking in Probabilistic Databases," proceedings of the VLDB endowment, vol. 2, no. 1, pp. 502-513, 2009 .

[5] V.-N. Huynh, Y. Nakamori and . J. Law, "A probability-based approach to comparison of fuzzy numbers and applications to target-oriented decision making," Fuzzy Systems, vol. 16, no. 2, pp. 371-387, 2008 .

[6] Available: https://www.tasnimnews.com/fa/news/1397.

با سپاس از شما