یادگیری‌ارجحیت فازی

داریوش حسن‌پور آده

دانشگاه صنعتی اصفهان، دانشکده مهندسی برق و کامپیوتر

Email: d.hasanpoor@ec.iut.ac.ir

چکیده—یادگیری‌ارجحیت یکی از زیررشته‌های یادگیری‌ماشین می‌باشد که هدف اصلی‌اش یادگیری ارجحیت‌های قابل پیش‌بینی از روی اطلاعات ارجحیت می‌باشد. از نقطه‌نظر یادگیری باناظر یادگیری‌ارجحیت روی یک دسته از عناصر که نسبت به یک دسته از برچسب‌ها یا عناصر ارجحیت بیشتری دارند آموزش داده می‌شوند که بتواند در نهایت مدل ارجحیت عناصر دیده نشده را پیش‌بینی کند. یادگیری‌ارجحیت فازی ترکیبی از یادگیری ماشین و سیستم‌های فازی می‌باشد که سعی در یادگیری ارجحیت عناصر از طریق یادگیری یک تابع رتبه‌بند با استفاده از منطق فازی می‌باشد.

کلمات‌کلیدی —*یادگیری ارجحیت، یادگیری ارجحیت فازی، تابع ارجحیت، طبقه‌بندی*

١ مقدمه

یادگیری ارجحیت[١] یکی از شاخه‌های یادگیری‌ماشین [٢] می باشد که در سال‌های اخیر توجه زیادی را به سمت خود جلب کرده است. وظیفه‌ی اصلی یادگیری‌ارجحیت، *یادگیری رتبه‌بندی کردن* عناصر می‌باشد. که با توجه به نوع اطلاعات آموزشی نوع ارجحیت و همچنین نوع رتبه‌بندی می‌تواند تغییر کند. در حالت کلی یادگیری‌ارجحیت به استنتاج کردن ارتباطات بین اعضای یک دسته و نگاشت این ارتباطات به مدلی که بتواند ارجحیت نسبی موجود بین اعضای این دسته بخوبی پیش‌بینی کند.

اگر بخواهیم تعریف دقیقی برای یادگیری‌ارجحیت ارائه دهیم می‌توانیم بگوییم که یادگیری‌ارجحیت به وظیفه‌ی یادگیری پیش‌بینی کردن رابطه‌ی ترتیب [٣] در مجموعه‌ای از اشیا گفت.[٢] یادگیری‌ارجحیت از داده‌های اطلاعات ارجحیت [۴] برای اهداف یادگیری خود استفاده می‌کند که این اطلاعات ارجحیت نقش مهمی در تصمیم‌گیری‌های خودکار در زمینه‌های متعددی مانند تئوری تصمیم‌گیری‌های کیفی[۵]، استدلال غیریکنواخت[۶]، ارضای محدودیت[٧] و طرح‌ریزی[٨] و ... کاربردهای فراوانی دارد. در مرحله ی آموزش، الگوریتم‌های یادگیری‌ارجحیت به داده‌های ترتیب رتبه بندی[٩] (نیمه)شناخته شده عناصر دسترسی دارند. بسته به مدل و نوع مساله الگوریتم‌های یادگیری‌ارجحیت می‌توانند به ٣ دسته‌ی رتبه‌بندی اشیا[١٠]، رتبه‌بندی برچسب‌ها[١١] و رتبه‌بندی نمونه‌ها[١٢] تقسیم‌بندی کرد.

در این نوشتار ابتدا مقدمه‌ای بر مفاهیم بنیادی مطرح در یادگیری‌ارجحیت آورده شده است و درنهایت با تمرکز بر مقاله اصلی سمینار شرح مفصلی در مورد یادگیری‌ارجحیت فازی ارائه خواهیم داد.

---

[١] Preference Learning
[٢] Machnine Learning
[٣] Order
[۴] Preference Information
[۵] Qualitative decision theory
[۶] Non-monotonic reasoning
[٧] Constraint satisfaction
[٨] Planning
[٩] Ranking Order
[١٠] Object Ranking
[١١] Label Ranking
[١٢] Instance Ranking

۲ یادگیری‌ارجحیت

برای اینکه بتوانیم *یادگیری‌ارجحیت فازی* را ارائه دهیم نیاز هست که در این قسمت به معرفی خلاصه‌ای از مفاهیم اولیه و بنیادی زمینه‌ی یادگیری‌ارجحیت بپردازیم.

۱.۲ *نمادگذاری*

طبق هر شاخه‌ی علمی دیگر به یک سری نماد برای پایه گذاری یادگیری‌ارجحیت نیاز داریم که در این قسمت به معرفی آن ها می‌پردازیم.

تعریف ۱.۲ (ارجحیت ضعیف): یک ارجحیت ضعیف با نماد $\succeq$ بروی مجموعه‌ای مانند $\mathcal{A}$ یک رابطه‌ی بازتابی و تعددی می‌باشد.

تعریف ۲.۲ (ارجحیت اکید): $a \succ b \leftrightarrow (a \succeq b) \wedge (b \not\succeq a)$

در راستای معناشناسی یادگیری‌ارجحیت تعاریف ۱.۲ و ۲.۲ می‌توان آن‌ها را به صورت جدول ۱ تعبیر کرد.

نماد	تعبیر
$a \succeq b$	"جایگزین a حداقل به اندازه‌ی جایگزین b ترجیح داده می‌شود."
$a \succ b$	"جایگزین a بیشتر از جایگزین b ترجیح داده می‌شود."

جدول ۱: تعبیر نمادهای روابط ارجحیت

تعریف ۳.۲ (ترتیب‌اکید کلی[۱۳] – رتبه‌بندی[۱۴]): اگر $\mathcal{A}$ یک مجموعه‌ای از اشیا / جایگزین‌ها[۱۵] $\{a_1, \ldots, a_m\}$ باشد، یک رتبه بندی از $\mathcal{A}$ یک جایگشتی همانند τ از مجموعه‌ی $\{1, \ldots, m\}$ می‌باشد بگونه‌ای که $a_i \succ a_j \leftrightarrow \tau(i) < \tau(j)$.

در تعریف ۳.۲ مقدار m تعداد اشیا/جایگزین‌ها و a_iها همان اشیا/جایگزین‌ها می‌باشند، که از این به بعد به مجموعه‌ی اشیا یا جایگزین‌ها، فقط جایگزین گفته خواهد شد. تعریف ۳.۲ در واقع فرمت[۱۶] خروجی کلی الگوریتم‌های یادگیری‌ارجحیت را ارائه می‌دهد، به گونه‌ای که خروجی ارجحیت ورودی‌ها یک جایگشتی اندیسی[۱۷] ترتیب‌اکید ارجحیت‌های ورودی‌ها می‌باشد بگونه‌ای که **اندیس** آن جایگزینی که **دارای ارجحیت بیشتری** نسبت به دیگری است، در مجموعه‌ی τ **مقدم‌تر** از دیگری می آید.

طبق آنچه که در تعریف ۳.۲ آمده، واضح است که الگوریتم‌های یادگیری‌ارجحیت درواقع یک جستجوکننده در فضای جایگشتی مجموعه τ می‌باشد؛ فضای جایگشت‌های τ را با $\mathcal{S}_m$ نمایش می‌دهند.

۲.۲ *انواع الگوریتم‌های رتبه‌بندی*

الگوریتم‌های رتبه‌بندی از نظر نوع وظیفه‌ای که به عهده دارند به ۳ تیپ تقسیم‌بندی می‌شوند که عبارتند از رتبه‌بندی برچسب‌ها، رتبه‌بندی اشیا و رتبه‌بندی نمونه‌ها؛ که در این قسمت شرح مختصری از این ۳ تیپ الگوریتم ارائه خواهیم داد.[۲]

۱.۲.۲ *رتبه‌بندی برچسب‌ها:*

در این رتبه‌بندی هدف این است که به ازای یک نمونه برچسب‌ها را براساس ارجحیت مرتب کند. شماتیک ورودی‌ها و خروجی‌ها این نوع الگوریتم در شکل ۱ آمده است.

Given:
- A set of training instances $\{x_k \mid k = 1, \ldots, n\} \subseteq \mathcal{X}$ (each instance typically represented by a feature vector)
- A set of labels $\mathcal{L} = \{\lambda_i \mid i = 1, \ldots, m\}$
- For each training instance x_k: a set of associated *pairwise preferences* of the form $\lambda_i \succ_{x_k} \lambda_j$

Find:
- A ranking function in the form of an $\mathcal{X} \longrightarrow \mathcal{S}_m$ mapping that assigns a ranking (permutation) $\succ_x$ of $\mathcal{L}$ to every $x \in \mathcal{X}$

شکل ۱: شماتیک کلی الگوریتم‌های رتبه‌بندی برچسب‌ها

همانطور که در شکل ۱ آمده است الگوریتم‌های رتبه‌بندی برچسب‌ها تعدادی نمونه‌ی آموزشی می‌گیرد که این نمونه‌های

[۱۳] Total Strict Order
[۱۴] Order
[۱۵] Alternatives
[۱۶] Format
[۱۷] Index Permutation

آموزشی معمولا به صورت بردار ویژگی‌ها نمایش داده می‌شود. سپس یک سری برچسب و همچنین به ازای هرکدام از نمونه‌ها یک مجموعه‌ی دوبه‌دو مرتب که ارجحیت برچسب‌ها برای هرکدام از نمونه‌ها مشخص شده است را به عنوان ورودی به الگوریتم داده می‌شود.الگوریتم‌های متعلق به رتبه‌بندی برچسب‌ها باید به عنوان خروجی یک تابع رتبه‌بند است که بتواند به ازای یک شی جایگشت های برچسب‌های متعلق به آن شی را با توجه به ارجحیتی که دارند برگرداند.

۲.۲.۲ *رتبه‌بندی اشیا:* در این رتبه‌بندی هدف این است که یک دسته از اشیا را برحسب ارجحیتی که نسبت به هم دارند، را مرتب‌سازی کنیم. که شماتیک ورودی‌ها و خروجی‌ها این نوع الگوریتم در شکل ۲ آمده است.

Given:

- A (potentially infinite) set $\mathcal{X}$ of objects (each object typically represented by a feature vector)
- A finite set of pairwise preferences $x_i \succ x_j$, $(x_i, x_j) \in \mathcal{X} \times \mathcal{X}$

Find:

- A ranking function $r(\cdot)$ that, given a set of objects $\mathcal{O} \subseteq \mathcal{X}$ as input, returns a permutation (ranking) of these objects

شکل ۲: شماتیک کلی الگوریتم‌های رتبه‌بندی اشیا

در شماتیکی که در شکل ۲ آمده است می‌بینیم که همانند رتبه‌بندی برچسب‌ها یک دسته از اشیا را به عنوان ورودی می‌گیرد. سپس یک دسته‌ی (تکه‌ای)مرتب از ارجحیت‌های نسبی این اشیا را نیز به عنوان ورودی به الگوریتم می‌دهیم و نهایت الگوریتم باید یک تابع رتبه‌بند را بیابد که بتواند ارجحیت‌های هر زیرمجموعه از مجموعه‌ی جهانی اشیا را در قالب یک جایگشتی از اشیا ورودی به عنوان خروجی برگرداند.

توجه شود که مجموعه‌ی (تکه‌ای)مرتب از ارجحیت اشیا که به عنوان داده‌آموزشی به الگوریتم داده می‌شود نیازی ندارد که تمامی ارجحیت‌های موجود میاد اشیا را شامل باشد، به چند دلیل این شرط می‌تواند به قابل پیاده‌سازی بودن الگوریتم کمک زیادی کند. اولین دلیل که بدیهی‌ترین دلیل می‌باشد، این است که ممکن است در کابردهای دنیای واقعی این الگوریتم تمامی ارجحیت های موجود بین اشیا نشاخته نشده باشد. دومین دلیل این است که در صورتی که تمامی ارجحیت های بین اشیا شناخته شده باشد اندازه ی مجموعه‌ی مرتب از ارجحیت خیلی بیشتر از تعداد اشیا می‌باشد؛ که از حل ۱ می‌بینیم این افزونگی داده‌ای تقریبا $\frac{n}{2}$ برابر تعداد اشیا می‌باشد که در کابردهای واقعی این الگوریتم‌ها معمولا تعداد اشیا زیاد است که در نتیجه تعداد ارتباطات مرتب بین آنها بسیار بیشتر از تعداد اشیا می‌باشد که مشکلات خواص خودش را در پی دارد که خارج از حیطه‌ی این نوشتار است.

$$\frac{\binom{n}{2}}{n} = \frac{n-1}{2} \tag{1}$$

در نتیجه این شرط که الگوریتم‌های رتبه‌بندی اشیا تنها با در اختیار داشتن مجموعه‌ی تکه‌ای مرتب از ارجحیت های میان اشیا باید بتواند مدلی برای رتبه‌بندی اشیا بدست بیاورد؛ باعث می شود که این الگوریتم‌های در کاربردهای دنیای واقعی کاربرد پیدا کند.

۳.۲.۲ *رتبه‌بندی نمونه:* رتبه‌بندی نمونه همانند رتبه‌بندی اشیا می‌باشد یعنی علاوه بر شماتیک معرقی شده در رتبه‌بندی اشیا ۲ عدد ورودی دیگر را نیز دارد. به این صورت که یک مجموعه اکیدا مرتب(تعریف ۲.۲ را ببینید) از برچسب‌ها را نیز اختیار می کند که در این مجموعه مرتب از برچسب‌ها آن برچسبی که دارای ارجحیت بیشتری است در ابتدای مجموعه ظاهر شود. و همچنین یک ورودی دیگر نیز به الگوریتم داده می‌شود که مجموعه ارتباطات یک‌به‌یک از هرکدام از اشیا را به برچسب متعلق به آن شی می باشد. همچنین در این نوع از الگوریتم‌ها نیازی به مجموعه (تکه ای)مرتب از ارجحیت‌های نسبی اشیا به الگوریتم داده شود.

۳.۲ *موارد خاص یادگیری‌ارجحیت*

همان‌طور که در قسمت ۲.۲ آمده است الگوریتم‌های یادگیری‌ارجحیت از نظر اهداف و کاربرد با یک‌دیگر متفاوت هستند، ما همه روزه در میان الگوریتم‌های طبقه‌بندی روزمره‌ای که با آنها سروکار دارم در واقع داریم از الگوریتم‌های ارجحیت استفاده می

کنیم که در اینجا به معرفی دو نوع از الگوریتم‌های طبقه‌بندی می پردازیم که حالت خاصی از الگوریتم‌های یادگیری‌ارجحیت می باشند.

۱.۳.۲ *طبقه‌بندی:* الگوریتم‌های طبقه‌بند[18] که همه روزه از آن‌ها استفاده می‌کنیم در واقع حالت خاصی از الگوریتم‌های رتبه بندی برچسب‌ها می‌باشد. از آنجایی که در طبقه‌بندی به ازای یک ورودی هدف پیش‌بینی برچسب مرتبط با آن ورودی می‌باشد، می توان داده‌ی ورودی را به یکی از الگوریتم‌های یادگیری‌ارجحیت (رتبه‌بندی برچسب‌ها) داد سپس از میان مجموعه جایگشت برچسب‌ها که به عنوان خروجی رتبه‌بند برگشت داده خواهد شد، آن برچسبی را که دارای بیشتری ارجحیت است را به عنوان برچسب منتخب برای آن نمونه ورودی انتصاب کرد. به عبارت دیگر می توان با استفاده از تعریف ۲ الگوریتم‌های طبقه‌بند را با استفاده از رتبه‌بند برچسب‌های یادگیری‌ارجحیت مدل کرد.

$$\mathcal{C}_{x_k} = \{\lambda_i \mid \lambda_i \succ_{x_k} \lambda_j,\ ۱ \leq j \neq i \leq m\} \qquad (۲)$$

۲.۳.۲ *طبقه‌بندی چند برچسب:* در طبقه‌بندی چند برچسب[19] هر نمونه به یک زیرمجموعه‌ای از برچسب‌ها نسبت داده‌می‌شود. این نوع از الگوریتم‌ها را نیز حالت خاصی از الگوریتم های یادگیری‌ارجحیت (رتبه‌بندی برچسب‌ها) می‌باشد به‌گونه‌ای که الگوریتم رتبه‌بندی برچسب‌ها یک عدد به عنوان تعداد برچسب های نسبت داده شده به اشیا داده می‌شود و الگوریتم بعد از مرتب‌سازی برچسب‌ها بر اساس ارجحیت‌هایی که دارند یک تعداد از برچسب‌ها که دارای ارجحیت بیشتری می‌باشند را به عنوان خروجی بر می‌گرداند. یا یک روش دیگر برای مدل کردن طبقه بند چند برچسب با استفاده از رتبه‌بند برچسب‌ها این است که در کنار داده‌های الگوریتم یک حد آستانه مابین [۰, ۱] به الگوریتم داده شود و بعد از رتبه‌بندی برچسب‌ها و نرمال‌سازی میزان ارجحیت های برچسب‌ها آن برچسب‌هایی که میزان ارجحیت آن‌ها بیشتر از حد آستانه به عنوان خروجی برگشت داده شود. در حالت کلی می توان طبقه‌بند چند برچسب را به صورت تعریف ۳ با استفاده از رتبه بند ارجحیت مدل کرد.

$$\mathcal{C}_{x_k} = \{\mathcal{L}_k \mid \lambda_i \succ_{x_k} \lambda_j,\ \lambda_i \in \mathcal{L}_k, \lambda_j \in \mathcal{L} \backslash \mathcal{L}_k\} \qquad (۳)$$

۴.۲ *انواع روش‌های یادگیری‌ارجحیت*

برای یادگیری‌ارجحیت دو روش معمول موجود است که یکی بر اساس یادگیری یک تابع سودمندی[20] و دیگری یادگیری ارتباطات ارجحیت[21] موجود بین داده‌ها را یاد می‌گیرید که برای درک بهتر روش ارائه شده در مقاله اصلی در این بخش به توضیح مختصری نسبت به هریک می‌پردازیم.

۱.۴.۲ *یادگیری تابع سودمندی:* یک راه طبیعی برای نشان دادن ارجحیت‌های موجود بین داده‌ها این‌است که جایگزین‌های موجود را با استفاده از یک تابع مطلوبیت ارزیابی کنیم. برای رتبه بندی اشیا یک تابع نگاشت $f : \mathcal{X} \to \mathcal{U}$ که مقدار سودمندی $f(x)$ به هریک از اشیا x تخصیص می‌دهد را یاد می‌گیریم و درنهایت اشیا را با معیار سودمندی‌ای که از این تابع سودمندی یادگرفته شده محاسبه می‌شود مرتب کرده و به عنوان خروجی برمی‌گردانیم. و در رتبه‌بندی برچسب‌ها به ازای هریک از برچسب‌ها $\lambda_i,\ i = ۱, \ldots, m$ یک تابع سودمندی $f_i : \mathcal{X} \to \mathcal{U}$ یادگرفته می‌شود و در نهایت برچسب‌ها را برحسب مقدار سودمندی‌ای که آن برچسب‌ها دارند مرتب می‌کنیم، $\lambda_i \succ_x \lambda_j \Rightarrow f_i(x) \geq f_j(x)$.

۵.۲ *یادگیری ارتباطات ارجحیت*

یک روش معمول دیگر در یادگیری‌ارجحیت این است که بیاییم و ارتباطات ارجحیت دودویی موجود بین دو اشیا(برچسب ها) را یاد بگیریم. برای رتبه‌بندی اشیا یک تابع ارجحیت دودویی مانند $\mathcal{Q}(x,\ x')$ یاد می‌گیرد که نشان می‌دهد آیا شی x نسبت به x'

[18] Classification
[19] Multi-label classification
[20] Learning Utility Functions
[21] Learning Preference Relations

ارجحیت دارد یا خیر. در نهایت ترتیب نهایی شامل یک جایگشتی می‌باشد که حداکثر سازگاری را با این ترتیب‌های دودویی داشته باشند. به عبارت دیگر جایگشتی را انتخاب می‌کنیم که مقدار ۴ را حداکثر کند.

$$V_{\mathcal{X}} = \sum_{i=۱}^{n} \sum_{j\neq i=۱}^{n} \mathcal{Q}(x_i, x_j) \qquad (۴)$$

برای رتبه‌بندی برچسب‌ها هم یک روشی به‌نام طبقه‌بندی دوبه‌دو ارائه شده‌است. ایده‌ی این طبقه‌بند دوبه‌دو این است که به ازای هریک از برچسب‌ها یک مدلی مانند $\mathcal{M}_{i,j}, ۱ \le i < j \le m, (\lambda_i, \lambda_j) \in \mathcal{L}$ را با استفاده از داده های آموزشی بدست بیاوریم(یاد بگیریم)؛ در نتیجه به تعداد $\frac{m(m-۱)}{۲}$ تعداد مدل نیاز داریم. داده‌های آموزشی شامل اطلاعات ارجحیت برچسب‌ها مانند $\lambda_i \succ_x \lambda_j$ می‌باشند که به‌صورت مثال های آموزشی (a, b) تبدیل می‌شوند که برای یادگیری مدل $\mathcal{M}_{a,\,b},\ a = \min(i,j),\ b = \max(i,j)$ استفاده می‌شوند. در واقع مدل $\mathcal{M}_{a,b}$ سعی بر یادگیری مدلی دارد که نگاشت ۵ را انجام دهد.

$$\mathcal{M}_{a,b} \mapsto \begin{cases} ۱ & \lambda_a \succ_x \lambda_b \\ ۰ & \lambda_b \succ_x \lambda_a \end{cases} \qquad (۵)$$

که این نگاشت می‌تواند توسط یکی از طبقه‌بند کننده‌های دودویی انجام گیرد.

۳ یادگیری‌ارجحیت فازی

مقاله اصلی[۱] یادگیری‌ارجحیت را با استفاده از انتگرال چوکت[۲۲] پیاده‌سازی کرده است. که در این قسمت بعد از یادآوری اندازه‌گیری غیرافزایشی[۲۳] و اینکه چرا انتگرال چوکت می‌تواند برای حل چنین مساله‌ای مفید باشد بحثی خواهیم داشت؛ سپس مروری بر انتگرال چوکت و در انتها روش اصلی ارائه شده در مقاله اصلی را بیان می‌کنیم. نتایج حاصله نیز در بخش بعدی این نوشتار به تفصیل آمده است.

[۲۲] Choquet Integral
[۲۳] Non-additive Measures

۳.۱ *اندازه‌گیری غیرافزایشی*

برای ارائه‌ی یک دید کاربردی از معنی و مفهوم اندازه گیری غیرافزایشی، در این قسمت سعی شده است که این ویژگی را متناسب با کاربردش در یادگیری‌ارجحیت با یک مثال فرضی ارائه دهیم.

فرض کنیم که یک مدرسه وجود دارد که به دانش مهندسی دانشجویان‌اش بیشتر از دانش ادبی آن‌ها اهمیت می‌دهد. وهمچنین در سیاست‌های این آموزشکده آمده است که بهتر است دانش آموختگانی پروش دهیم که در تمامی علوم آموزش داده‌شده، دارای دانش متعادلی باشند. حال اگر فرض کنیم تنها ۳ درس *ریاضی*، *فیزیک* و *ادبیات* در این آموزشکده آموزش داده می‌شود طبق اولین سیاست این آموزشکده که به دانش مهندسی بیشتر از دانش ادبی بها می‌دهد، به نمرات ریاضی و فیزیک دانشجویان وزن بیشتری می‌دهد ولی از طرف دیگر طبق سیاست دوم به دانشجویانی که نمرات متعادل‌تری نسبت به دیگری که واریانس نمرات آن‌ها زیاد است اهمیت نیز می‌دهد.

حال اگر بخواهیم طبق این سیاست‌ها یک اندازه‌گیر تعریف کنیم، مثلا می‌توانیم اندازه‌های زیر را ارائه دهیم که با جزییات مساله سازگارند.

$$\mu(\{\emptyset\}) = ۰ \qquad \mu(\{۱,\ ۲,\ ۳\}) = ۱,$$
$$\mu(\{۳\}) = ۰/۳ \qquad \mu(\{۱\}) = \mu(\{۲\}) = ۰/۴۵,$$
$$\mu(\{۱,\ ۲\}) = ۰/۵ \qquad \mu(\{۱,\ ۳\}) = \mu(\{۲,\ ۳\}) = ۰/۹,$$

که در اینجا طبق آنچه که در اندازه‌گیری‌های بالا مطرح شده است هر یک از دروس ریاضی، فیزیک و ادبیات به تریب با اعداد ۱ و ۲ و ۳ نمایش داده می‌شوند، و می‌توان این تفصیر را کرد که دانشجوی در تمامی دروس خود خوب عمل می‌کند دارای حداکثر درجه‌ی اهمیت یعنی مقدار ۱ برای آموزشکده می‌باشد و دانشجویانی که در هیچ یک از دروس خوب عمل نمی‌کند اهمیتی برای گروه آموزشی آموزشکده ندارند و کسب نمره‌ی خوب در هریک از دروس ریاضی یا فیزیک به تنهایی دارای درجه‌ی اهمیت یکسان ۰.۴۵ می‌باشد این مقدار بیشتر از مقدار درجه‌ی اهمیتی است که برای درس ادبیات

گذاشته که مقدار ۰.۳ تخصیص داده شده است می‌باشد. و در انتها دانشجویانی که در کنار یکی از دروس ریاضی یا فیزیک، در درس ادبیات نیز نمره‌ی خوبی کسب کرده‌باشد دارای درجه‌ی اهمیت بیشتری نسبت به دانشجویانی که فقط در درس ریاضی یا فیزیک نمره خوبی کسب کرده‌اند می‌باشد. که در اندازه‌گیری‌های ارائه شده برای این مثال می‌توان ارتباط ۶ را مشاهده می‌کنیم.

$$\mu(\{1,2\}) \neq \mu(\{1\}) + \mu(\{2\}) \tag{6}$$

که رابطه‌ی ۶ در واقع تعریف اندازه‌گیری‌های غیرافزایشی می‌باشد. که در حالت کلی برای اندازه‌گیری‌های که خواص ۷...۹ را دارند، اندازه‌گیری‌های غیرافزایشی می‌گویند.

$$\mu(\emptyset) = 0,\ \mu(\mathcal{X}) = 1 \tag{7}$$

$$\mu(A) \leq \mu(B) \quad \forall A \subseteq B \subseteq \mathcal{X} \tag{8}$$

$$\mu(\{a_1,\ \dots,\ a_k\}) \neq \sum_{i=1}^{k} \mu(\{a_i\}) \tag{9}$$

یکی از روش‌های مفید ارائه شده برای نمایش اندازه‌گیری های غیرافزایشی که در مقاله کمک زیاد برای یادگیری مدل ارجحیت کرده است نگاشت‌موبیوس[24] می‌باشد که به صورت ۱۰ تعریف می‌شود.

$$\mu(\mathcal{B}) = \sum_{A \subseteq \mathcal{B}} m(A) \tag{10}$$

به ازای تمامی $\mathcal{B} \subseteq \mathcal{X}$ که نگاشت‌موبیوس $m_{\hat{\mu}}$ را که اندازه‌ی $\hat{\mu}$ به صورت ۱۱ تعریف می‌شود.

$$m_{\hat{\mu}}(A) = \sum_{v \subseteq A} (-1)^{|A|-|v|} \hat{\mu}(v) \tag{11}$$

که در واقع مقدار بدست آمده از ۱۱ در واقع می‌توان به عنوان وزن ارزشی‌ای است که اختصاصا به مجموعه‌ی A داده‌شده است و ارزش ارتباطات غیرمستقیم اجزای تشکیل دهنده‌ی آن مجموعه حذف گردیده‌است.

آنچه که در مثال ابتدای این بخش دیدیم در واقع یک مساله‌ی یادگیری‌ارجحیت می‌باشد، که دیدم ارجحیت‌های این مساله یک اندازه‌گیری غیرافزایشی می‌باشد. در اکثر مسائل یادگیری ارجحیت نوع ارتباطات میان ارجحیت‌ها از نوع اندازه‌گیری‌های غیرافزایشی می‌باشد. به همین علت و طبق آنچه که در مورد انتگرال چوکت مشهور است که به‌خوبی می‌تواند با اندازه‌گیری های غیرافزایشی کار کند منطقی به نظر می‌رسد که ترکیب این دو مساله، یعنی یادگیری‌ارجحیت و انتگرال چوکت می‌تواند منجر به نتیجه‌ای خوب شود. که همین دلیل انگیزه‌ی اصلی مطرح شده در مقاله نسبت به اینکه چرا از انتگرال‌فازی‌چوکت برای یادگیری ارجحیت استفاده شده است.

۲.۳ *اهمیت معیارها*

اندازه‌گیری‌های افزایشی را می‌توان به صورت زیر نمایش داد،

$$\mu(A) = \sum_{x_i \in A} \mu(\{x_i\}) = \sum_{x_i \in A} w_i$$

که اندازه‌ی گیری $w_i = \mu(\{x_i\})$ وزن معیار[25] x_i می‌باشد. طبق آنچه که در ۷ و ۸ آمده است این وزن‌ها غیر صفر و $\sum_{i=1}^{n} w_i = 1$ می‌باشند. که در اینجا آنچه که واضح است این است که میان معیار ها(x_iها) هیچ ارتباطی وجود ندارد؛ یعنی اندازه‌گیری روی معیار x_i مستقل از نتیجه‌ی اندازه‌گیری انجام شده روی معیار x_j می باشد.

حس ضرورت اندازه‌گیری اهمیت معیارها زمانی پدیدار می شود که اندازه‌گیری ما از نوع غیرافزایشی باشد. حال فرض کنید یک اندازه‌گیر فازی μ بروی مجموعه‌ای از معیارها مانند X داده شده است. میزان اهمیت معیار x_i به صورت میانگین میزان افزایش اهمیت با اضافه کرده معیار x_i به یک مجموعه‌ی دیگر مانند $A \subset X \setminus \{x_i\}$ که این میانگین‌گیری را می‌توان توسط رابطه‌ی ۱۲ نمایش داد.

$$\varphi(x_i) = \frac{1}{n} \sum_{A \subset X \setminus \{x_i\}} \frac{1}{\binom{n-1}{|A|}} (\mu(A \cup \{x_i\}) - \mu(A)) \tag{12}$$

که میزان اهمیت معیارها روی یک اندازه‌گیر مانند μ توسط یک برداری همانند $\varphi(\mu, X) =< \varphi(x_1),\ \dots,\ \varphi(x_n) >$ نشان داده

[24] Mobius Transform

[25] Criteria

می‌شود که دارای خواص ۱۳ و ۱۴ می‌باشد.

$$0 \le \varphi(x_i) \le 1 \tag{13}$$

$$\sum_{i=1}^{n} \varphi(x_i) = 1 \tag{14}$$

۳.۳ *انتگرال‌چوکت*

در این قسمت مروری خلاصه بر انتگرال‌چوکت که مورد استفاده در مقاله‌ی اصلی صورت گرفته است، می‌کنیم و سپس با استفاده از نگاشت موبیوس که در بخش ۳.۱ تعریف شد تبدیلی رو انتگرال‌چوگت اعمال می‌کنیم که در نهایت مساله‌ی یادگیری ارجحیت به یک مساله‌ی بهینه‌سازی تبدیل شود[۲۶].

همان‌طور که می‌دانیم رابطه‌ی انتگرال‌چوکت به صورت ۱۵ بیان می‌شود[۳].

$$\begin{aligned} \mathcal{C}_\mu(f) &= \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)})) \cdot \sum_{k=i}^{n} \mu(\{x_{(k)}\}) \\ &= \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)})) \cdot \mu(A_{(i)}) \end{aligned} \tag{15}$$

بگونه‌ای که $(\cdot)$ یک جایگشتی از مجموعه‌ی $\mathcal{S}_m$ است بگونه‌ای که $0 \le f(x_{(1)}) \le f(x_{(2)}) \le \ldots \le f(x_{(n)})$ که $f(x_{(0)}) = 0$ تعریف شده است و $A_{(i)} = \{x_{(i)},\ \ldots,\ x_{(n)}\}$.

حال اگر بخواهیم انتگرال‌چوکت را از نظر ظاهری با آنچه که برای میزان اهمیت معیارها در بخش ۳.۲ گفته شد براحتی قابل اثبات است که تابع انتگرال‌چوکت را می‌توان به صورت ۱۶ بازنویسی کرد.

$$\begin{aligned} \mathcal{C}_\mu(f) &= \sum_{i=1}^{n} (f(x_{(i)}) - f(x_{(i-1)})) \cdot \mu(A_{(i)}) \\ &= \sum_{i=1}^{n} f(x_{(i)}) \cdot \Big(\mu(A_{(i)}) - \mu(A_{(i+1)})\Big) \end{aligned} \tag{16}$$

حال اگر نگاشت‌موبیوس را بر معادله‌ی بازنویسی شده چوکت در ۱۶ اعمال کنیم به روابط ۱۷ می‌رسیم.

$$\mathcal{C}_\mu(f) = \sum_{i=1}^{n} f(x_{(i)}) \cdot \Big(\mu(A_{(i)}) - \mu(A_{(i+1)})\Big) \tag{17}$$

$$= \sum_{i=1}^{n} f(x_{(i)}) \cdot \sum_{R \subseteq \mathcal{T}} m(R)$$

$$= \sum_{T \subseteq \mathcal{X}} m(T) \times \min_{x_i \in T} f(x_i)$$

$$= \sum_{T \subseteq \mathcal{X}} \sum_{v \subseteq T} (-1)^{|A|-|v|} \mu(v) \times \min_{x_i \in T} f(x_i) \tag{18}$$

$$\mathcal{T} = \{\mathcal{G} \cup \{x_{(i)}\} \mid \mathcal{G} \subset \{x_{(i+1)},\ \ldots,\ x_{(n)}\}\} \tag{19}$$

که در ۱۸ تنها چیزی که نامعلوم است مقدار μ می‌باشد که در بخش بعدی توضیح داده خواهد شد که چگونه یادگرفته خواهد شد و می‌توان به صورت بردار ضرب داخلی ۲۰ نوشت.

$$\Big\langle m_\varphi,\ \varphi(f(x))\Big\rangle \tag{20}$$

که نگاشت $\varphi : \mathbb{R}^n \to \mathbb{R}^{2^n-1}$ به صورت ۲۱ تعریف می کنیم.

$$\begin{aligned} \varphi(x) &= \varphi(x_1,\ \ldots,\ x_2) \\ &= \Big(x_1,\ \ldots,\ x_n,\ \min\{x_1,\ x_2\},\ \ldots,\ \min\{x_{n-1},\ x_n\}, \\ &\qquad \min\{x_1,\ x_2,\ x_3\},\ \ldots,\ \min\{x_1,\ \ldots,\ x_n\}\Big) \end{aligned} \tag{21}$$

که m_φ ماتریس $\Big\langle m_1,\ \ldots,\ m_n,\ m_{n+1},\ \ldots,\ m_{2^n-1}\Big\rangle$ مقادیر بدست آمده توسط نگاشت‌موبیوس می‌باشد که ترتیبش توسط $\varphi(x)$ تعیین می‌شود.

۴.۳ *یادگیری‌ارجحیت با استفاده از انتگرال‌چوکت*

ایده این مقاله این است که برای رتبه‌بندی اشیا که در بخش ۲.۲.۲ معرفی شد، بیاییم تابع رتبه‌بند را در قالب انتگرال‌چوکت ارائه دهیم. اگر فرض کنیم که هریک از اشیا مانند $o \in \mathcal{O}$ به صورت یک برداری از ویژگی‌ها نمایش دهیم.

$$f_o = \Big(f_o(x_1),\ \ldots,\ f_o(x_n)\Big)$$

[۲۶] بنابه اینکه انتگرال‌چوکت جز سرفصل‌های تدریس شده در کلاس بوده و کاملا با مفهوم این انتگرال‌فازی آشنا هستیم از توضیح راجع به جزییات بنیادی این انتگرال خودداری می‌کنم.

که مقدار $f_o(x_i)$ مقدار ارزیابی شده شی o براساس معیار x_i می باشد. رتبه‌ی شی o را می‌توان با استفاده از انتگرال‌چوکت که در ۱۷ و ۱۸ معرفی شد به صورت زیر می‌تواند باشد.

$$U(o) = \mathcal{C}_\mu(f_o) \quad (۲۲)$$

حال فرض کنیم داده‌های آموزشی به صورت مجموعه‌ای از اشیا(برداری از ویژگی‌های منتصب به هر شی) به همراه برچسب l_i هر کدام از داده‌ها می‌باشد، $i = ۱, \ldots, N$. از این داده‌ها یک داده‌ی جدید مانند $\mathcal{D}$ استخراج می‌شود که رابطه‌ی ارجحیت اشیا داخل دیتاست اولیه را به صورت دودویی مدل می‌کند به‌طوری که، $(o_i, o_j) \in \mathcal{D}$ که نشان می‌دهد که $o_i \succ o_j$ به شرطی که اگر در دیتاست اولیه داشته باشیم (o_i, l_i) و (o_j, l_j) به‌طوری که $l_i > l_j$.
در راستای ایده‌ی به حداقل رساندن ریسک بصورت تجربی[۲۷]، مقاله سعی کرده تعداد خطاهای رتبه‌بندی را با استفاده از انتگرال چوکت بر روی مجموعه داده‌آموزشی $\mathcal{D}$ به حداقل برساند. از آن جایی که انتگرال‌چوکت منحصرا توسط معیار μ بر روی مجموعه ی $X = \{x_۱, \ldots, x_n\}$ مشخص می‌شود، بنابراین تعریف مناسب این معیار اهمیت زیادی دارد – که به همین علت بود که از نگاشت موبیوس در ۱۷ استفاده کرده است.

مقاله با الهام از اصل‌حداکثرحاشیه[۲۸] در روش‌های مبتنی بر کرنل[۲۹] یادگیری‌ماشین مساله‌ی یادگیری μ را به عنوان یک مساله‌ی بهینه‌سازی مدل کرده است.

$$\max_{M,\, \xi_۱,\, \ldots,\, \xi_۲} \left\{ M - \frac{\gamma}{|\mathcal{D}|} \sum_{(o_s,\, o_t) \in \mathcal{D}} \xi_s + \xi_t \right\} \quad (۲۳)$$

به‌طوری که دارای شروط تعریف شده از ۲۴ تا ۲۸ می‌باشد. که مقدار M نشان دهنده‌ی حاشیه‌ای است که می‌خواهیم حداکثراش کند – که می‌شود حداقل اختلاف میان میزان سودمندی دو شی که مانند o_s و o_t می‌باشد به‌گونه‌ای که $o_s \succ o_t$. به‌طور خاص‌تر اگر بخواهیم بگوییم M یک حاشیه‌ی نرم است که با احتساب این حقیقت که درحالت کلی ارضای تمام نابرابری‌ها به طور هم‌زمان

غیرممکن است در نتیجه هر شی o_s با یک متغییر ضعیف[۳۰] همانند ξ_s در ارتباط است. که این متغییرهای ضعیف نامنفی هستند و مقدار مثبت این متغییر جریمه‌ای متناسب با اندازه خودش تحمیل متحمل می‌شود. و در نهایت γ یک مقدار موازنه[۳۱] می‌باشد که هرچقدر مقدارش بیشتر باشد میزان جریمه‌ی تحمیل شده از سوی متغییرهای ضعیف بیشتر می‌باشد.

$$\mathcal{C}_\mu(f_{o_s}) - \mathcal{C}_\mu(f_{o_t}) > M - \xi_s - \xi_t \quad \forall (o_s,\, o_t) \in \mathcal{D} \quad (۲۴)$$

$$\xi_s \geq ۰ \quad (۲۵)$$

$$\sum_{T \subseteq X} m(T) = ۱ \quad (۲۶)$$

$$\sum_{B \subseteq A} m(B) \geq ۰ \qquad \forall A \subseteq X \quad (۲۷)$$

$$\sum_{L \subseteq A} m(L) \leq \sum_{K \subseteq B} m(K) \qquad \forall A \subset B \subseteq X \quad (۲۸)$$

بعد از این‌که مقاله با استفاده از معادلات اثبات می‌کند که با استفاده از انتگرال‌چوکت مساله‌ی یادگیری‌ارجحیت‌فازی به یک مساله‌ی بهینه‌سازی تبدیل می‌شود، توضیحی در مورد الگوریتم بهینه‌سازی مورد استفاده در آزمایشات خود نداده است.

۴ آزمایشات

۱.۴ *داده‌ها*

در این مقاله ۱۵ عدد دیتاست برای آزمایشات مورد استفاده قرار گرفته‌اند که اکثر آن‌ها از UCI و از چارچوب WEKA جمع آوری شده اند. که همگی آن‌ها معیارهای سنجش[۳۲] برای آزمودن الگوریتم‌های یادگیری می‌باشند. علاوه بر این‌ها، مقاله تعدادي مجموعه داده از دنياي واقعي، از منابع ديگر مانند داده‌هايي از يک فرآيند پلي‌استر صنعتي رنگرزي و اطلاعات در مورد ارزيابي مجلات رياضي جمع آوري کرده است. جدول ۲ خلاصه اي از تمام مجموعه‌داده‌ها را نشان مي‌دهد که در سايت نويسندگان[۳۳] مقاله در دسترس عموم گذاشته‌اند.

[۲۷] Empirical Risk Minimization
[۲۸] Maximum Margin Principle
[۲۹] Kernel-Based
[۳۰] Slack
[۳۱] Trade-Off
[۳۲] Bench-mark
[۳۳] http://www.uni-marburg.de/fb12/kebi/research

data set	#instances	#attributes	#classes	source
Color(CLR) 1-7	120	3	3	[5]
Scientific Journals(SCJ)	172	5	4	[6]
CPU	209	6	2	UCI
Auto MPG	398	8	6	UCI
Employee Selection(ESL)	488	4	9	WEKA
Mamographic(MMG)	830	5	2	UCI
Lecture Evaluation(LEV)	1000	4	5	WEKA
Concrete Compressive Strength(CCS)	1030	8	6	UCI
Car Evaluation(CEV)	1728	6	4	UCI

جدول ۲: دیتاست‌ها و ویژگی‌های آن‌ها

در ادامه، شرح مختصري در مورد ۲ دیتاست از مجموعه داده‌های جدول ۲ آورده شده است.

۱.۱.۴ *دیتاست CPU:* این یک مجموعه‌ی داده، معیار سنجش استاندارد از مخزن UCI است. که در رابطه با عملکرد نسبی CPUهای می‌باشد که شامل ۱۰ عدد ویژگي، که سه تا از آن آنجايي که به وضوح فاقد ارزش پیش گویي بوده اند حذف شده‌اند(نام‌فروشنده، نام‌مدل، EPR[۳۴]). که مشخصات ویژگی‌های مورد استفاده از این دیتاست در جدول ۴ آمده است. در جدول ۵ نیز چند سطر از داده‌های این دیتاست آمده است که در این دیتاست برچسب(ویژگی PRP) دارای مقداری پیوسته می‌باشد.

۲.۱.۴ *دیتاست Mamographic:* این مجموعه‌ی داده، دارای اطلاعاتی در مورد سرطان پستان توسط که ماموگرافي بدست آمده است. هدف پیش‌بیني شدت(خوش‌خیم یا بدخیم) توده با استفاده از ویژگي هاي BI-RADS می‌باشد. این دیتاست دارای ۶ عدد ویژگی بوده که از این ۶ عدد ویژگی یکی از آن‌ها ویژگی هدف(خوش‌خیم یا بدخیم می‌باشد) و یکی از ویژگی‌ها قابل پیش بینی نمی‌باشد و دارای ۴ ویژگی دیگر که قابل پیش‌بینی می‌باشند. که مشخصات ویژگی‌های مورد استفاده از این دیتاست در جدول ۶ آمده است. در جدول ۷ نیز چند سطر از داده‌های این دیتاست آمده است.

[۳۴] Estimated Relative Performance

۲.۴ *مقایسه*

مقاله روش ارائه شده‌ی خود را برای رتبه‌بندی با کرنل های خطی و چندجمله‌ای روش RankSVM که یکی روش‌های مدرن[۳۵] برای مساله رتبه‌بندی می‌باشد مقایسه کرده است. که نتیجه‌ی این مقایسه‌ها در جدول ۳ آمده است. در این جدول که میزان دقت(همراه با انحراف‌معیار چندین اجرا) ۳ الگوریتم میانگین وزنی[۳۶] و RankSVM با ۳ درجه‌ی چندجمله‌ای کرنل و روش ارائه شده در مقاله(CI[۳۷]) آورده شده است. در کنار هرکدام از دقت ها عددی که داخل پرانتز آورده شده است رتبه‌ی دقت الگوریتم مربوطه در آن دیتاست را نمایش می‌دهد و در سطر آخر نیز میانگین رتبه‌ی هرکدام از الگوریتم‌ها آورده شده است. همان‌طور که آورده شده است روش ارائه شده با بیشترین میانگین دقت نتایج مطلوب تری را بدست داده است.

برای ملموس‌تر شدن نتایج بدست آمده در جدول ۳ جدول ۸ را که به‌نوعی ماتریس درهم‌ریختگی رتبه‌ی روش‌ها در اجرا بر روی دیتاست‌های مختلف ارائه شده است که درک بهتری نسبت به اینکه روش ارائه شده در مقاله دیگر روش‌های مدرن در این زمینه را مغلوب کرده است.

[۳۵] State of the art
[۳۶] Weighted Mean
[۳۷] Choquet Integral

data set	WM	PL d=1	PL d=2	PL d=3	CI
CLR 1	.9663±.0148(4)	.9506±.0155(5)	.9674±.0129(3)	.9700±.0141(2)	.9828±.0090(1)
CLR 2	.8740±.0293(4)	.8601±.0294(5)	.8876±.0200(3)	.9341±.0244(2)	.9804±.0128(1)
CLR 3	.9343±.0204(4)	.9268±.0219(5)	.9375±.0156(3)	.9633±.0143(2)	.9878±.0150(1)
CLR 4	.9357±.0171(4)	.9228±.0247(5)	.9431±.0189(3)	.9659±.0166(2)	.9915±.0056(1)
CLR 5	.9518±.0194(3)	.9485±.0179(5)	.9565±.0142(2)	.9516±.0171(4)	.9682±.0140(1)
CLR 6	.9046±.0202(4)	.8923±.0205(5)	.9127±.0201(3)	.9460±.0191(2)	.9825±.0121(1)
CLR 7	.8880±.0312(4)	.8797±.0256(5)	.8892±.0219(3)	.9258±.0237(2)	.9688±.0167(1)
SCJ	.8168±.0105(4)	.8098±.0112(5)	.8270±.0241(3)	.8313±.0109(2)	.8450±.0201(1)
CPU	.9965±.0027(3)	.9950±.0093(5)	.9978±.0012(2)	.9955±.0005(4)	.9986±.0014(1)
MPG	.8887±.0176(4)	.8850±.0143(5)	.8912±.0078(3)	.8967±.0093(2)	.9060±.0111(1)
ESL	.9497±.0162(2)	.9559±.0071(1)	.9465±.0104(4)	.9491±.0126(3)	.9424±.0098(5)
MMG	.8961±.0230(2)	.8536±.0168(4)	.8714±.0181(3)	.7813±.0350(5)	.9015±.0210(1)
LEV	.8710±.0289(2)	.8620±.0320(3)	.8713±.0250(1)	.8527±.0300(5)	.8610±.0320(4)
CCS	.8650±.0068(4)	.8586±.0102(5)	.8862±.0184(3)	.8962±.0203(2)	.9050±.0038(1)
CEV	.8981±.0066(4)	.8804±.0076(5)	.9118±.0059(3)	.9585±.0090(2)	.9771±.0039(1)
average rank	3.47	4.53	2.8	2.73	1.47

جدول۳: مقایسه‌ی دقت روش‌های متفاوت با روش RankSVM با کرنل‌های درجات ۱ تا ۳ و همچنین روش معمول میانگین وزنی

	WM	PL d=1	PL d=2	PL d=3	CI
WM	–	14	2	5	2
PL d=1	1	–	1	3	2
PL d=2	13	14	–	4	2
PL d=3	10	12	11	–	1
CI	13	13	13	14	–

جدول۸: ماتریس درهم‌ریختگی رتبه‌ی روش‌ها در اجرا بر روی دیتاست‌های مختلف ارائه شده در جدول ۲

همان‌طور که در جدول ۸ مشاهده می‌شود از ۱۵ دیتاست مورد استفاده در آزمایشات مقاله روش ارائه شده دارای بیشترین امتیاز برد نسبت به سایر الگوریتم‌ها مورد مقایسه می‌باشد که الگوریتم‌های مورد مقایسه از روش‌های مدرن در این زمینه می باشد.

۵ نتیجه‌گیری

در این مقاله کاربرد انتگرال چوکت را در مسائل یادگیری ارجحیت ارائه داده است. درواقع از انتگرال چوکت به عنوان تابع سودمندی برای مسائل رتبه‌بندی استفاده کرده است، که این انتگرال را بخاطر یک سری ویژگی‌های منحصر به فرد انتگرال چوکت که پیشتر آورده شده است انتخاب شده است؛ که مهم‌ترین این ویژگی ها این بود که این انتگرال می‌تواند ارتباط وابستگی ارزشی بین معیارها را که به طور ذاتی دارای خاصیت غیر افزایشی می‌باشند، به‌خوبی نشان دهد.

این مقاله نشان داده است که با بکارگیری انتگرال چوکت و نگاشت‌موبیوس مساله‌ی یادگیری‌ارجحیت فازی به یک مساله ی بهینه‌سازی تبدیل می‌شود که می‌توان با هر الگوریتم بهینه‌سازی اقدام به حل این مساله کرد.[۳۸] در نهایت با مقایسه‌ای که با مدرن ترین الگوریتم‌های موجود در این زمینه نشان داده است که این روش بهتر از بقیه نتیجه داده است.

مراجع

[1] A. F. Tehrani, W. Cheng, and E. Hullermeier, *"Preference Learning Using the Choquet Integral: The Case of Multipartite Ranking"*. in Fuzzy Systems. IEEE Transactions, Dec. 2012, vol. 20, pp. 1102–1113.

[2] J. Furnkranz and E. Hullermeier, *"Encyclopedia of Machine Learning"*. Springer, 2010, ch. Preference Learning, pp. 789–795.

[3] M. Safayani, *"Fuzzy integral lecture"*. In "Fuzzy Set and Systems' Course", Isfahan University Of Technology, Spring 2015.

[4] J. Furnkranz and E. Hullermeier, *"Preference Learning"*. Kunstliche Intelligenz, pp. 60–61, 2005.

[5] M. Nasiri and S. Berlik, *"Modeling of polyester dyeing using an evolutionary fuzzy system"*. in Proc. Joint Int. Fuzzy Syst. Assoc. World Congr. Eur. Soc. Fuzzy Logic Technol. Conf., 2009, pp. 1246–1251.

[6] G. Beliakov and S. James, *"Citation-based journal ranks: The use of fuzzy measures"*. Fuzzy Sets Syst., vol. 167, no. 1, pp. 101–119, 2011.

[۳۸] که البته مقاله در مورد الگوریتم بهینه‌سازی اعمال شده برای مدل ارائه شده در مقاله سخنی نگفته است.

Feature	Description	Data Type
MYCT	Machine cycle time in nanoseconds	Integer
MMIN	Minimum main memory in kilobytes	Integer
MMAX	Maximum main memory in kilobytes	Integer
CACH	Cache memory in kilobytes	Integer
CHMIN	Minimum channels in units	Integer
CHMAX	Maximum channels in units	Integer
PRP	Published relative performance(class)	Integer

جدول۴: ویژگی‌های مورد استفاده‌ی دیتاست CPU

Row#	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	PRP(Class)
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
3	29	8000	32000	32	8	32	220
4	29	8000	32000	32	8	32	172

جدول۵: چند مثال از دیتاست CPU

Feature	Description	Data Type
BI-RADS	Assessment, (non-predictive!)	Ordinal
Age	Patient's age in years	Integer
Shape	Mass shape	Nominal
Margin	Mass margin	Nominal
Density	Mass density	Ordinal
Severity	Intensity(class)	Binominal

جدول۶: ویژگی‌های مورد استفاده‌ی دیتاست Mamographic

Row#	BI-RADS	Age	Shape	Margin	Density	Severity(class)
1	5	67	3	5	3	1
2	4	43	1	1	?	1
3	5	58	4	5	3	1
4	4	28	1	1	3	0

جدول۷: چند مثال از دیتاست Mamographic