

# ارائه یک روش انتخاب ویژگی بر اساس مجموعه‌های فازی مردد برای یادگیری چندبرچسبی

امین هاشمی و محمدباقر دولتشاهی

گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه لرستان، خرم آباد، ایران

تاریخ پذیرش: ۱۴۰۰/۸/۱۴

تاریخ دریافت: ۱۴۰۰/۵/۵

## چکیده

امروزه روش‌های مبتنی بر انتخاب ویژگی با توجه به افزایش داده‌های حجیم مورد توجه قرار گرفته‌اند. علاوه بر این با تولید مداوم داده‌های مختلف و نیز ابعاد بالای این داده‌ها، روش‌های مؤثر در کاهش ابعاد به ویژه انتخاب ویژگی مورد نیاز هستند. بسیاری از داده‌ها را می‌توان در گروه داده‌های چندبرچسبی قرار داد. به این معنا که هر نمونه در مجموعه داده می‌تواند متعلق به بیش از یک کلاس داده باشد. در این مقاله، یک روش انتخاب ویژگی مبتنی بر مجموعه‌های فازی مردد برای کاهش ابعاد داده‌های چندبرچسبی مبتنی بر مفهوم مجموعه‌های فازی مردد ارائه شده است. در این روش ما از ترکیب سه معیار مختلف در اندازه‌گیری همبستگی بین ویژگی‌ها و برچسب‌ها و همچنین سه معیار شباهت برای اندازه‌گیری شباهت بین ویژگی‌ها استفاده کرده‌ایم. در واقع هر کدام از این روش‌ها را به عنوان یک خبره برای تصمیم‌گیری در انتخاب ویژگی در نظر گرفته‌ایم. ترکیب روش‌های همبستگی و شباهت بر اساس مفهوم انرژی اطلاعات در مجموعه‌های فازی مردد انجام گرفته است. برای نمایش مؤثر بودن روش پیشنهادی، مقایساتی با روش‌های جدید در حوزه انتخاب ویژگی چندبرچسبی صورت گرفته است. این مقایسات بر اساس معیارهای دقت، زیان همینگ و زمان اجرای الگوریتم انجام شده‌اند.

عبارات و کلمات کلیدی: مجموعه‌های فازی مردد، انتخاب ویژگی، داده‌های چندبرچسبی، داده‌های حجیم، انرژی اطلاعات

Email(s): hashemi.aml@fe.lu.ac.ir, dowlatshahi.mb@lu.ac.ir

۱۴۰۰ انجمن سیستم‌های فازی ایران

Mathematics Subject Classification (2010): 90C70

## ۱ مقدمه

در جهان امروز پژوهشگران با برخی مسائل مبهم روبرو می‌شوند که باعث می‌شود درجه‌ای از عدم قطعیت به یک مسئله وارد شود. برای حل این مشکل، زاده مجموعه‌های فازی<sup>۱</sup> [۲۵] را معرفی کرد. مجموعه‌های فازی امروزه بین پژوهشگران بسیار محبوب شده‌اند به شکلی که گسترش‌های بسیاری از آن‌ها از جمله مجموعه‌های فازی شهودی<sup>۲</sup> [۲]، نوع  $n$  [۲۲]، مجموعه‌های فازی مردد<sup>۳</sup> [۲۳] و غیره. توررا<sup>۴</sup> [۲۳] یک نسخه از مجموعه‌های فازی را به نام مجموعه‌های فازی مردد ارائه کرد. مزیتی که این روش به مجموعه‌های فازی اضافه کرد این است که نظرات چندین خبره نسبت به عدم قطعیت از طریق مجموعه‌های فازی مردد ارائه می‌شود. در نتیجه با استفاده از مجموعه‌های فازی مردد، پژوهشگران می‌توانند جنبه‌های مختلفی از عدم قطعیت را در یک مسئله در نظر بگیرند [۶].

امروزه دسته‌بندی چندبرچسبی<sup>۵</sup> در بسیاری از موارد دنیای واقعی از جمله بایوانفورماتیک، دسته‌بندی تصویر و دسته‌بندی متن به شکل گسترده‌ای مورد استفاده قرار می‌گیرد. داده‌های چندبرچسبی نوعی از داده‌ها هستند که در آن هر نمونه آموزشی متعلق به بیش از یک برچسب کلاس است. به فرآیند پیش‌بینی برچسب‌های یک نمونه جدید، دسته‌بندی چندبرچسبی گفته می‌شود. به عنوان مثال تصویر یک منظره را در نظر بگیرید که به صورت همزمان می‌تواند سه برچسب کوه، درخت و گل را داشته باشد. بنابراین استفاده از دسته‌بندی چندبرچسبی برای این گونه از داده‌ها بسیار کار دشواری است، چرا که روش‌های دسته‌بندی تک‌برچسب با انواعی از داده‌ها کار می‌کنند که در آنها هر نمونه آموزشی فقط متعلق به یک برچسب کلاس است. داده‌های چندبرچسبی اغلب شامل تعداد بسیاری ویژگی هستند که بخشی از این ویژگی‌ها اغلب نامرتب و یا تکراری است. وجود این گونه ویژگی‌ها در داده سبب بروز مشکلاتی از جمله هزینه بالای محاسبات، بیش برآزش<sup>۶</sup>، دقت دسته‌بندی پایین و زمان یادگیری بالا می‌شود

<sup>1</sup>Fuzzy sets

<sup>2</sup>Intuitionistic fuzzy sets

<sup>3</sup>Hesitant fuzzy sets

<sup>4</sup>Torra

<sup>5</sup>Multi-label classification

<sup>6</sup>Over-fitting

[۱۱]، [۱۴]، [۹]. انتخاب ویژگی<sup>۷</sup> یک روش مؤثر برای مقاله با مشکلات یاد شده است. در فرآیند انتخاب ویژگی، ویژگی‌های مرتبط در داده انتخاب می‌شوند و ابعاد داده با حذف ویژگی‌های غیرمرتبط و تکراری کاهش می‌یابد [۲۴]، [۱۰].

روش‌های انتخاب ویژگی چندبرچسبی بر اساس دو استراتژی مختلف مختلف شکل می‌گیرند: تحول دودویی<sup>۸</sup> و انطباق الگوریتم<sup>۹</sup>. روش‌های مبتنی بر رویکرد تحول دودویی، با مسئله انتخاب ویژگی چندبرچسبی مانند چندین مسئله انتخاب ویژگی تک‌برچسبی مستقل رفتار می‌کنند که در آن به ازای هر برچسب یک بار روش انتخاب ویژگی تک‌برچسبی اعمال می‌شود. بزرگترین عیب این رویکرد به این مسئله برمی‌گردد که در یک داده چندبرچسبی، برچسب‌های کلاس با هم همبستگی دارند که این همبستگی به طور کامل در رویکردهای تحول دودویی نادیده گرفته می‌شوند. بنابراین برای رفع این مشکل، رویکردهای مبتنی بر انطباق الگوریتم ارائه شده‌اند که در آنها یک الگوریتم به صورت اختصاصی فقط برای انتخاب ویژگی چندبرچسبی به کار گرفته می‌شود [۱۷]، [۲۰].

در یک دسته‌بندی دیگر، روش‌های انتخاب ویژگی چندبرچسبی را می‌توان بر اساس تعامل آنها با دسته‌بند<sup>۱۰</sup> به سه دسته فیلتر<sup>۱۱</sup>، پنهان<sup>۱۲</sup> و جاسازی شده<sup>۱۳</sup> تقسیم کرد [۱]. روش‌های فیلتر، ویژگی‌ها را قبل از فرآیند یادگیری به صورت کاملاً مستقل از دسته‌بند و بر اساس معیارهای از پیش تعیین شده ارزیابی می‌کنند. از طرفی روش‌های پنهان در نقطه مقابل روش‌های فیلتر قرار دارند و زیرمجموعه‌های مختلف ویژگی را بر اساس دقت بدست آمده از دسته‌بند می‌سنجند. در واقع مبنای انتخاب ویژگی در این روش‌ها فقط دسته‌بند است. در پایان در روش‌های جاسازی شده، یادگیری مدل فرآیند و انتخاب ویژگی به صورت همزمان صورت می‌گیرد. در واقع این روش‌ها نیز به مشابه روش‌های پنهان از دسته‌بند استفاده می‌کنند ولی با این تفاوت که برخلاف آنها به صورت تکراری از دسته‌بند استفاده نمی‌کنند و فقط یک بار ساخته می‌شود، که از دسته‌بند ساخته

<sup>7</sup> Feature selection

<sup>8</sup> Binary transformation

<sup>9</sup> Algorithm adaptation

<sup>10</sup> Classifier

<sup>11</sup> Filter

<sup>12</sup> Wrapper

<sup>13</sup> Embedded

شده برای ارزیابی ویژگی‌ها استفاده می‌شود. [۴]، [۲۸].

امروزه، بسیاری از داده‌ها از نوع چندبرچسبی هستند. این دسته از داده‌ها با افزایش تعداد برچسب‌ها، چالش‌های فراوانی را برای الگوریتم‌های یادگیری دارند. با توجه به دلایلی که در ابتدای این بخش در مورد ضرورت انتخاب ویژگی در فرآیند یادگیری و نیز در نظر گرفتن این مسئله که روش‌های انتخاب ویژگی حال حاضر، همچنان برای داده‌های چندبرچسبی به دقت دسته‌بندی قابل قبولی دست نیافته‌اند، بنابراین ارائه روش‌های انتخاب ویژگی موثر یکی از چالش‌های اساسی در داده‌های چندبرچسبی است [۱۳]، [۱۸]. در کنار این روش‌ها، تکنیک‌های شورایی<sup>۱۴</sup> در مسائل یادگیری ماشین عملکرد مؤثری داشته‌اند. این تکنیک‌ها در ارزیابی ویژگی‌ها نیز بسیار مؤثر هستند، چرا که هر روش یا معیار ارزیابی ویژگی فقط جنبه‌های خاصی از ویژگی‌ها را در نظر می‌گیرد. در رویکرد شورایی به جای یک روش، ترکیبی از روش‌ها و یا معیارهای انتخاب ویژگی در نظر گرفته می‌شود [۳]، [۸]، [۱۵].

در این مقاله، ما روشی را ارائه کرده‌ایم که به دنبال حداکثرسازی همبستگی بین ویژگی‌ها و برچسب‌های کلاس و حداقل‌سازی افزونگی بین ویژگی‌ها است. در واقع در این روش ویژگی‌ها از دو منظر همبستگی و افزونگی ارزیابی می‌شوند و ترکیب آنها برای انتخاب زیرمجموعه نهایی ویژگی‌ها مورد استفاده قرار می‌گیرد. در هر دو حالت ما از یک رویکرد شورایی استفاده کرده‌ایم. به شکلی که به صورت شورایی با استفاده از سه روش مختلف همبستگی بین ویژگی‌ها و برچسب‌ها محاسبه می‌شود و با استفاده از سه روش دیگر نیز شباهت بین ویژگی‌ها اندازه‌گیری می‌شود. ترکیب این روش‌ها بر اساس مفهوم مجموعه‌های فازی مردد صورت گرفته است. به طور کلی روش ارائه شده در مقاله از یک استراتژی سه مرحله‌ای استفاده می‌کند: در مرحله اول، روش‌های همبستگی با برچسب‌ها و در مرحله دوم شباهت بین ویژگی‌ها محاسبه می‌شود. در این دو مرحله روش‌ها با استفاده از مدل‌سازی بر اساس مجموعه‌های فازی مردد و مفهوم انرژی اطلاعات در این مجموعه‌ها ترکیب می‌شوند. در پایان نتایج بدست آمده در دو مرحله با استفاده از یک معیار شایستگی<sup>۱۵</sup> با هم ترکیب می‌شوند و ویژگی‌ها بر اساس این معیار رتبه‌بندی می‌شوند. قابل ذکر است که این الگوریتم بر اساس یک رویکرد فیلتر و بر اساس استراتژی

<sup>14</sup>Ensemble

<sup>15</sup>Merit

انطباق الگوریتم در نظر گرفته شده است. برای اثبات عملکرد روش ارائه شده، این روش را با پنج الگوریتم جدید و قدرتمند در زمینه انتخاب ویژگی چندبرچسبی که بر اساس رویکرد فیلتر هستند، مقایسه کرده‌ایم. این مقایسات بر اساس سه معیار دقت دسته‌بندی، زیان همینگ<sup>۱۶</sup> و زمان اجرا و بر روی پنج مجموعه داده چندبرچسبی صورت گرفته است. نتایج نشان از برتری روش ارائه شده نسبت به روش‌های مقایسه شده دارد.

## ۲ کارهای مرتبط

در طول چند سال اخیر تلاش‌های بسیاری در زمینه انتخاب ویژگی چندبرچسبی توسط پژوهشگران انجام شده است که در این بخش به معرفی چند مورد از این روش‌ها که بر اساس استراتژی‌های کاملاً متفاوتی طراحی شده‌اند، می‌پردازیم. روش PPT-Relief یک روش انتخاب ویژگی چندبرچسبی منطبق بر استراتژی تحول دودویی است. این روش قبل از اجرای انتخاب ویژگی برای کاهش پیچیدگی محاسباتی مسئله ناشی از اجرای تکراری الگوریتم بر اساس تعداد برچسب‌ها، مجموعه داده را هرس می‌کند. به این صورت که برچسب‌هایی که در کمتر از ده نمونه آموزشی ظاهر شده‌اند از مجموعه داده حذف می‌شوند. پس از این کار، معیار Relief که از روش‌های فیلتر در محاسبه همبستگی بین یک ویژگی و یک برچسب است، بر روی داده اجرا می‌شود و در نهایت نتایج با هم ترکیب می‌شوند [۲۱]. پنیری و همکاران [۱۹] یک روش مبتنی بر الگوریتم بهینه‌سازی کلونی مورچگان ارائه کردند. در این روش ویژگی‌ها به یک گراف کامل مدل‌سازی شدند و از عملکرد مورچه‌ها در این گراف کامل برای ارزیابی ویژگی‌ها استفاده شده است. هاشمی و همکاران [۱۲] برای اولین بار یک مسئله انتخاب ویژگی را به یک مسئله تصمیم‌گیری چندشاخه مدل کردند. در این مدل‌سازی که برای انتخاب ویژگی چندبرچسبی صورت گرفت، ویژگی‌ها به عنوان کاندیداها و برچسب‌ها به عنوان خبره‌ها در سیستم تصمیم‌گیری چندشاخه در نظر گرفته شدند. در این روش ابتدا ماتریس تصمیم با استفاده از همبستگی بین ویژگی‌ها و برچسب‌ها محاسبه و سپس وزن خبره‌ها بر

<sup>16</sup>Hamming loss

یک روش انتخاب ویژگی بر اساس مجموعه‌های فازی مردد برای یادگیری — ۲۰۶

اساس انتروپی<sup>۱۷</sup> در نظر گرفته شد. در پایان از الگوریتم تاپسیس<sup>۱۸</sup> برای رتبه‌دهی به ویژگی‌ها استفاده شده است. کاشف و همکاران [۱۶] نیز یک روش انتخاب ویژگی را به نام ClusterPareto برای داده‌های چندبرچسبی ارائه کردند. این روش از مفهوم غلبه پارتو<sup>۱۹</sup> و مرتبط‌سازی غیرمسلط<sup>۲۰</sup> برای پیدا کردن ویژگی‌های غیرمغلوب استفاده کرده است. هم‌چنین برای رتبه‌دهی به ویژگی‌های غیرمغلوب نیز از یک روش خوشه‌بندی استفاده شده است. روش LRFS [۲۷] یک روش انتخاب ویژگی چندبرچسبی دیگر است که با استفاده از معیار اطلاعات متقابل شرطی<sup>۲۱</sup> به ویژگی‌ها رتبه می‌دهد. هم‌چنین این روش افزونگی بین برچسب‌ها را نیز در نظر می‌گیرد.

### ۳ مفاهیم پایه

در این بخش به بررسی مفاهیم پایه مورد استفاده در روش پیشنهادی می‌پردازیم. دسته‌بندی چندبرچسبی، مجموعه‌های فازی مردد، الگوریتم رتبه‌بندی، معیارهای شباهت و یک مقدار شایستگی در این بخش مرور می‌شوند.

#### ۱.۳ دسته‌بندی چندبرچسبی

در یک داده چندبرچسبی، هر نمونه آموزشی دارای یک بردار ویژگی و یک بردار دودویی برچسب است. در واقع هر نمونه آموزشی می‌تواند متعلق به بیش از یک برچسب کلاس باشد. دسته‌بندی چندبرچسبی به فرآیندی گفته می‌شود که در آن یک مدل از نمونه‌های آموزشی ایجاد می‌شود که از طریق آن می‌توان پیش‌بینی کرد که یک نمونه جدید متعلق به کدام برچسب‌های کلاس است [۱۷].

---

<sup>17</sup>Entropy

<sup>18</sup>TOPSIS

<sup>19</sup>Pareto

<sup>20</sup>Non-dominated sorting

<sup>21</sup>Conditional mutual information

### ۲.۳ مجموعه‌های فازی مردد

**تعریف ۱.۳** (عنصر فازی مردد). اگر  $X$  را يك مجموعه ثابت در نظر بگیریم، آنگاه مجموعه فازی مردد  $A$  بر روی  $X$  با شرط تابع  $h_A(x)$  تعریف می‌شود. خروجی این تابع مقداری در بازه  $[0, 1]$  است. مجموعه  $A$  به صورت زیر تعریف می‌شود [۵]:

$$A = \{ \langle x, h_A(x) \rangle \mid x \in X \} \quad (1)$$

که در آن  $h_A(x)$  مجموعه تمام مقادیر ممکن در بازه  $[0, 1]$  است. در واقع  $h_A(x)$ ، عنصر فازی مردد <sup>۲۲</sup> نامیده می‌شود [۲۳].

**تعریف ۲.۳** (حد بالا و پایین). اگر يك عنصر فازی مردد داشته باشیم، آنگاه حد بالا و پایین عنصر به این صورت تعریف می‌شوند:

$$h^-(x) = \min h(x) : \text{حد پایین}$$

$$h^+(x) = \max h(x) : \text{حد بالا}$$

**تعریف ۳.۳** (انرژی اطلاعات). برای يك مجموعه فازی مردد  $A = \{ \langle x, h_A(x) \rangle \mid x_i \in X, i = 1, 2, \dots, n \}$ ، انرژی اطلاعات  $A$  به صورت زیر تعریف می‌شود:

$$E(A) = \sum_{i=1}^n \left( \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A\sigma(j)}^2(x_i) \right) \quad (2)$$

که  $n$  برابر کاردینالیتی <sup>۲۳</sup> مجموعه ثابت،  $l_i$  برابر با تعداد مقادیر عضویت <sup>۲۴</sup> و  $h_{A\sigma(j)}(x_i)$ ،  $j$ -امین عنصر از  $i$ -امین مجموعه ثابت مجموعه فازی مردد مربوطه است [۵].

**تعریف ۴.۳** (همبستگی بین مجموعه‌های فازی مردد). ضریب همبستگی بین دو مجموعه

<sup>22</sup> Hesitant fuzzy element

<sup>23</sup> Cardinality

<sup>24</sup> Membership values

فازی مردد  $A$  و  $B$  به صورت زیر تعریف می‌شود [۵]:

$$C(A, B) = \sum_{i=1}^n \left( \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A\sigma(j)}(x_i) \cdot h_{B\sigma(j)}(x_i) \right) \quad (۳)$$

که پارامترهای آن همان پارامترهای تعریف ۳.۳ هستند.

تعریف ۵.۳ (ضریب همبستگی بین مجموعه‌های فازی مردد). ضریب همبستگی<sup>۲۵</sup> بین دو مجموعه فازی مردد  $A$  و  $B$  به صورت زیر تعریف می‌شود [۵]:

$$\rho(A, B) = \frac{C(A, B)}{[C(A, A)]^{\frac{1}{2}} \cdot [C(B, B)]^{\frac{1}{2}}} \quad (۴)$$

$$= \frac{\sum_{i=1}^n \left( \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A\sigma(j)}(x_i) \cdot h_{B\sigma(j)}(x_i) \right)}{\left[ \sum_{i=1}^n \left( \frac{1}{l_i} \sum_{j=1}^{l_i} h_{A\sigma(j)}^2(x_i) \right) \right]^{\frac{1}{2}} \cdot \left[ \sum_{i=1}^n \left( \frac{1}{l_i} \sum_{j=1}^{l_i} h_{B\sigma(j)}^2(x_i) \right) \right]^{\frac{1}{2}}}$$

که باید شرایط زیر را ارضا کند:

$$0.1 \quad \rho(A, B) = \rho(B, A)$$

$$0.2 \quad 0 \leq \rho(A, B) \leq 1$$

$$0.3 \quad \text{اگر } A = B \text{ باشد، آنگاه } \rho(A, B) = 1$$

شرط اول بیانگر این است که ماتریس ضریب همبستگی باید متقارن باشد. شرط دوم بر این موضوع اشاره دارد که ضرایب همبستگی باید مقداری در بازه  $[0, 1]$  داشته باشند. در نهایت، شرط سوم بیانگر این است که حداکثر مقدار ضرایب همبستگی زمانی بدست می‌آید که دو مجموعه فازی مردد دقیقاً یکسان باشند.

### ۳.۳ الگوریتم‌های محاسبه همبستگی بین ویژگی‌ها و برجسب‌ها

در این بخش به معرفی سه معیار رتبه‌بندی ویژگی‌ها بر اساس همبستگی آنها با برجسب‌های کلاس می‌پردازیم. این معیارها به صورت زیر تعریف می‌شوند:

<sup>25</sup>Correlation coefficient



### ۱- ضریب همبستگی پیرسون

ضریب همبستگی پیرسون<sup>۲۶</sup> که برای محاسبه میزان مرتبط بودن یک ویژگی با یک برچسب استفاده می‌شود. مقدار ضریب همبستگی پیرسون بین یک ویژگی و برچسب در بازه [۰, ۱] قرار دارد، به شکلی که هر چه قدر مقدار به یک نزدیک‌تر باشد، همبستگی بین آنها بیشتر می‌شود. ضریب همبستگی پیرسون به صورت زیر برای ویژگی  $p$  و برچسب  $q$  محاسبه می‌شود:

$$\begin{aligned} \text{Correlation}(p, q) &= \frac{\text{Cov}(p, q)}{\sqrt{\text{var}(p) \cdot \text{var}(q)}} \quad (۵) \\ &= \frac{\sum_{i=1}^n (p_i - \bar{p}) \cdot (q_i - \bar{q})}{\sqrt{(\sum_{i=1}^n p_i - \bar{p})^2 \cdot (\sum_{i=1}^n q_i - \bar{q})^2}} \end{aligned}$$

که  $\text{Cov}(p, q)$  برابر کواریانس ویژگی  $p$  و برچسب  $q$ ،  $\text{var}(p)$  و  $\text{var}(q)$  هم واریانس آنها است. این رابطه برای هر ویژگی در مقابل همه برچسب‌ها محاسبه می‌شود تا در نهایت ماتریس همبستگی به ابعاد  $p \times q$  بدست آید. همچنین  $n$  برابر تعداد نمونه‌های آموزشی،  $p_i$  و  $q_i$  برابر مقدار ویژگی و برچسب در نمونه  $i$ -ام،  $\bar{p}$  و  $\bar{q}$  نیز میانگین مقادیر ویژگی  $p$  و برچسب  $q$  هستند [۱۳].

### ۲- شباهت کسینوسی

شباهت کسینوسی بین ویژگی  $p$  و برچسب  $q$  از طریق ضرب داخلی بردارهای آنها بدست می‌آید. در واقع زاویه نزدیکی دو بردار از نظر جهت‌گیری و نه از نظر اندازه را به صورت زیر محاسبه می‌کند:

$$\text{CosineSimilarity}(p, q) = \cos(\theta) = \frac{\sum_{i=1}^n (p_i \cdot q_i)}{\sqrt{\sum_{i=1}^n p_i^2 \times \sum_{i=1}^n q_i^2}} \quad (۶)$$

پارامترهای این رابطه نیز مانند روش قبلی است و مقدار شباهت کسینوسی در بازه [۰, ۱] قرار می‌گیرد. مقدار شباهت کسینوسی صفر به این معناست که در زاویه ۹۰ درجه با هم قرار دارند و با هم شباهتی ندارند. هر چه قدر مقدار شباهت کسینوسی به یک نزدیک‌تر باشد، زاویه بین بردارهای آنها کوچک می‌شود و در نتیجه همبستگی بالاتر می‌رود [۱۹].

### ۳- رگرسیون خط الرأس

<sup>26</sup> Pearson Correlation Coefficient

یک روش انتخاب ویژگی بر اساس مجموعه‌های فازی مردد برای یادگیری — ۲۱۰

رگرسیون خط الرأس<sup>۲۷</sup> یک روش رایج برای تنظیم کردن<sup>۲۸</sup> در مسائل خطاهای حداقل مربعات خطی<sup>۲۹</sup> که به دنبال کاهش عواقب چندخطی بودن<sup>۳۰</sup> در رگرسیون می‌باشد. فرض کنید  $X \in \mathbb{R}^{N \times M}$  ماتریس ویژگی‌ها با  $N$  نمونه و  $M$  ویژگی است. همچنین  $Y \in \mathbb{R}^{N \times L}$  ماتریس برجسب‌ها با  $N$  نمونه آموزشی و  $L$  برجسب است. حال  $W \in \mathbb{R}^{M \times L}$  یک ماتریس ضرایب است که رابطه بین ویژگی‌ها و برجسب‌ها را نشان می‌دهد. این ماتریس به صورت زیر محاسبه می‌شود:

$$W = \operatorname{argmax}_w (\|Y - XW\|_F^2 + \lambda \|W\|_F^2) = (XX^T + \lambda I)^{-1} X^T Y \quad (۷)$$

که  $I \in \mathbb{R}^{N \times N}$  یک ماتریس همانی و  $\lambda$  یک پارامتر مثبت باشد که ضرایب را به گونه‌ای منظم کند که اگر مقدار زیادی گرفته باشند، تابع بهینه‌سازی جریمه شود. لازم به ذکر است که ماتریس ضرایب  $W$  از طریق داده آموزشی و با رگرسیون خط الرأس تخمین زده می‌شود که می‌تواند میزان تمایز ویژگی‌ها را اندازه‌گیری کند. هرچه قدر مقدار  $W_{ij}$  بزرگ باشد، بیانگر این است که ویژگی  $i$ -ام، همبستگی بالایی با برجسب  $j$ -ام دارد. در این مقاله ما مقدار  $\lambda$  را به صورت تجربی برابر ۱۰ در نظر گرفته‌ایم. با توجه به این که مقادیر ضرایب همبستگی پیرسون و شباهت کسینوسی در بازه  $[0, 1]$  قرار دارند، برای یکسان‌سازی مقادیر، ماتریس  $W$  را هم در بازه  $[0, 1]$  قرار می‌دهیم و از طریق رابطه زیر نرمال می‌کنیم:

$$\operatorname{Ridge}(i, j) = \frac{W_{ij} - \min W_i}{\max W_i - \min W_i} \quad (۸)$$

که  $W_{ij}$  برابر مقدار همبستگی ویژگی  $i$ -ام نسبت به برجسب  $j$ -ام بر اساس ماتریس  $W$ ،  $\min W_i$  برابر کمترین و  $\max W_i$  برابر بزرگترین مقدار ویژگی  $i$ -ام است. نهایت، ماتریس نرمال  $\operatorname{Ridge}$  بدست می‌آید. با توجه به این که ماتریس‌های بدست آمده برای امتیازدهی به ویژگی‌ها باید به بردار تبدیل شوند، در نتیجه در این مقاله ما حداکثر مقدار هر ویژگی را در برابر همه برجسب‌ها به عنوان امتیاز نهایی آن ویژگی در هر سه

<sup>27</sup>Ridge Regression

<sup>28</sup>Regularize

<sup>29</sup>Linear least-squares problems

<sup>30</sup>Multicollinearity

الگوریتم قرار داده‌ایم [۱۲].

### ۴.۳ معیارهای شباهت

در این بخش ۳ روش محاسبه شباهت بین ویژگی‌ها که در روش پیشنهادی مورد استفاده قرار گرفته‌اند، معرفی می‌شوند. اولین معیار که برای محاسبه شباهت بین ویژگی‌های  $p$  - $m$  و  $q$  - $m$  استفاده می‌شود، معکوس فاصله اقلیدسی<sup>۳۱</sup> ( $IED$ ) است که به صورت زیر محاسبه می‌شود [۵]:

$$IED(p, q) = \frac{1}{Euclidean(p, q)} = \frac{1}{\|p - q\|_2} \quad (9)$$

دومین معیار شباهت، قدرمطلق ضریب همبستگی پیرسون است. این مقدار بین دو ویژگی  $p$  و  $q$  بر اساس رابطه ۵ محاسبه می‌شود. در نهایت معیار سوم نیز شباهت کسینوسی است که در رابطه ۶ تعریف شده است.

### ۵.۳ معیار شایستگی مبتنی بر همبستگی

در این بخش یک معیار شایستگی معروف در زمینه انتخاب ویژگی که توسط هال [۷] معرفی شده است را بررسی می‌کنیم. آنها رابطه‌ای را معرفی کردند که از طریق آن همبستگی بین ویژگی‌ها و برجسب کلاس و هم افزونگی بین ویژگی‌ها به صورت همزمان در نظر گرفته می‌شود. این رابطه به صورت زیر است:

$$merit = \frac{K \times \overline{RCF}}{\sqrt{K + K(K - 1) \times \overline{RFF}}} \quad (10)$$

بر اساس رابطه بالا،  $K$  بیانگر تعداد ویژگی‌ها است و  $\overline{RCF}$  نمایانگر وابستگی بین ویژگی‌ها و برجسب کلاس آنها و  $\overline{RFF}$  نیز افزونگی بین ویژگی‌ها است. با توجه به رابطه بالا می‌توان دریافت که شایستگی زمانی بیشترین مقدار خود را دارد که وابستگی بین ویژگی و برجسب کلاس حداکثر و شباهت بین ویژگی‌ها حداقل باشد.

<sup>31</sup>Inverse Euclidean Distance

## ۴ روش پیشنهادی

در این بخش روش پیشنهادی به نام  $ML - HFS^{32}$  ارائه شده است. این روش شامل ۳ بخش اساسی است. بخش اول افزونگی بین ویژگی‌ها را با تشکیل مجموعه‌های فازی مردد و بر اساس معیارهای شباهت در نظر می‌گیرد. در بخش دوم، همبستگی بین ویژگی‌ها و برجسب‌های کلاس بر اساس الگوریتم‌های رتبه‌دهی و مفهوم مجموعه‌های فازی مردد محاسبه می‌شود. در قسمت سوم، بر اساس مقادیر بدست آمده توسط در دو مرحله قبل برای هر ویژگی، زیرمجموعه دلخواه از ویژگی‌ها بر اساس معیار شایستگی مبتنی بر همبستگی بدست می‌آید. مراحل الگوریتم ارائه شده به صورت مرحله به مرحله به صورت زیر است:

گام اول) محاسبه همبستگی بین ویژگی‌ها و برجسب‌ها بر اساس مجموعه‌های فازی مردد:

بر اساس مفاهیم پایه فازی مردد، یک بردار مقادیر عضویت (به عنوان مثال برداری از نظرات خبره‌ها) برای تشکیل  $HFS$  ها مورد نیاز است. بنابراین در این ۳ حالت، بردار بدست‌آمده از حداکثر مقادیر در روش‌های ضریب همبستگی پیرسون، شباهت کسینوسی و رگرسیون خط الرأس به عنوان خبره‌ها در این مسئله در نظر گرفته می‌شوند که هر کدام سه بردار هستند. برای این منظور، ویژگی‌ها ابتدا بر اساس این سه خبره، رتبه‌دهی می‌شوند و بردارهای رتبه به عنوان نظرات خبره‌ها در نظر گرفته می‌شوند. بنابراین اگر  $d$  ویژگی داشته باشیم، باید  $d$  بردار  $RHFS$  (مجموعه‌های فازی مبتنی بر رابطه بین ویژگی‌ها و برجسب‌ها) تشکیل شود و این بردارها بر اساس رابطه زیر محاسبه می‌شوند:

$$RHFS_m = \{ \langle F_m, h_{R_m}(F_m) \rangle \mid F_m \in F : m = 1, 2, \dots, d \} \quad (11)$$

که  $RHFS_m$  برابر  $m$ -امین مجموعه فازی مردد بر اساس امتیاز بدست‌آمده برای ویژگی شماره  $m$  است. لازم به ذکر است که در مجموعه‌های فازی بر خلاف مجموعه‌های واضح، برای هر عضو یک تابع عضویت آن عضو را در مجموعه نشان می‌دهد. در رابطه بالا نیز این تابع  $h_{R_m}$  است.  $F_m$  برابر  $m$ -امین ویژگی و  $F$

<sup>32</sup>Multi-Label Hesitant Fuzzy Set

مجموعه همه ویژگی‌ها  $F = \{F_1, F_2, \dots, F_m, \dots, F_d\}$  است. از طرفی  $h_{R_m}(F_m)$  برابر است با:

$$(۱۲)$$

$$h_{R_m}(F_m) = \{Correlation(F_m), CosineSimilarity(F_m), Ridge(F_m)\}$$

که مقادیر  $Correlation(F_m)$ ،  $CosineSimilarity(F_m)$  و  $Ridge(F_m)$  امتیازهای اختصاص یافته توسط الگوریتم‌های همبستگی بین ویژگی‌ها و برجسب‌ها به ویژگی شماره  $m$  هستند. سپس، انرژی اطلاعات برای هر ویژگی از طریق رابطه زیر بدست می‌آید:

$$E_{RHFS_m} = \frac{1}{3} \sum_{j=1}^3 h_{RHFS_m}^j(F_m) \quad (۱۳)$$

$E_{RHFS_m}$  هر ویژگی در بردار  $Rel(m)$  که بیانگر نتایج همبستگی ویژگی‌ها و برجسب‌ها است، قرار می‌گیرد.

گام دوم) محاسبه شباهت بین ویژگی‌ها بر اساس مجموعه‌های فازی مردد:

روال این گام هم دقیقاً مشابه گام اول است. ابتدا سه ماتریس معکوس فاصله اقلدیسی  $IED$ ، شباهت کسینوسی  $CS$  و ضریب همبستگی پیرسون  $PC$  محاسبه می‌شوند. مانند حالت قبل سه بردار برای هر ماتریس با توجه به حداکثر مقدار هر ویژگی در هر ماتریس بدست می‌آید. در نهایت مجموعه فازی شباهت  $SHFS$  برای هر ویژگی به صورت زیر محاسبه می‌شود:

$$SHFS_m = \{ \langle x_m, h_{S_m}(x_m) \rangle \mid x_m \in X \} \quad (۱۴)$$

$$h_{SHFS}(x_m) = \{IED_m, PC_m, CS_m\} \quad (۱۵)$$

که  $X$  یک مجموعه ثابت است و  $x_m$  بیانگر مقدار افزونگی (شباهت) ویژگی  $m$  در برابر سایر ویژگی‌ها است. بعد از ایجاد  $SHFS$  ها بر اساس مقادیر ۳ بردار بدست‌آمده

یک روش انتخاب ویژگی بر اساس مجموعه‌های فازی مردد برای یادگیری — ۲۱۴

از معیارهای شباهت، انرژی اطلاعات هر ویژگی را بر اساس افزونگی به صورت زیر محاسبه می‌کنیم:

$$E_{HFS}(SHFS_m) = \frac{1}{3} \sum_{j=1}^3 h_{SHFS_m}^j(x_m) = \frac{1}{3} (IED_m^j, PC_m^j, CS_m^j) \quad (16)$$

که حداکثر مقدار در ماتریس بدست‌آمده برای هر ویژگی در بردار  $Red(m)$  که بیانگر میزان شباهت بین ویژگی‌ها است، قرار می‌گیرد.

گام سوم) در این گام شایستگی هر ویژگی بر اساس رابطه ۱۰ محاسبه می‌شود که به صورت زیر بدست می‌آید:

$$merit = \frac{d \times Rel}{\sqrt{d + d(d-1) \times Red}} \quad (17)$$

که  $d$  تعداد ویژگی‌ها است.

در پایان ویژگی‌ها بر اساس مقدار شایستگی بدست‌آمده به صورت نزولی مرتب می‌شوند و کاربر می‌تواند به میزان دلخواه از آنها انتخاب کند.

## ۵ آزمایش‌ها و نتایج

در این بخش عملکرد روش پیشنهادی با پنج روش جدید و قدرتمند در زمینه انتخاب ویژگی چندبرچسبی مقایسه می‌شود. این روش‌ها که همگی از نوع فیلتر هستند عبارتند از: ClusterPareto [۱۶]، LRFS [۲۷]، MFS-MCDM [۲۷]، PPT-ReliefF، [۲۱] و MLACO [۱۹].

### ۱.۵ مجموعه داده‌ها

برای اندازه‌گیری عملکرد روش‌های مختلف، ۵ مجموعه داده دنیای واقعی برگرفته از مخزن مولان<sup>۳۳</sup> استفاده شده است. استفاده شده است. این مجموعه داده‌ها از نظر تعداد

<sup>33</sup><http://mulan.sourceforge.net/datasets-mlc.html>

جدول ۱: مشخصات مجموعه داده‌های چندبرچسبی

مجموعه داده	تعداد نمونه‌ها	تعداد ویژگی‌ها	تعداد برچسب‌ها
Arts	۷۴۸۴	۴۶۲	۲۶
Bibtex	۷۳۹۵	۱۸۳۶	۱۵۹
Chemistry	۶۹۶۱	۵۴۰	۱۷۵
Recreation	۵۰۰۰	۶۰۶	۲۲
Science	۵۰۰۰	۷۴۳	۴۰

ویژگی و برچسب متنوع هستند. بنابراین برای این که نمایش دهیم روش پیشنهادی برای مجموعه‌های داده با ابعاد متفاوت عملکرد مناسبی دارد، این مجموعه‌های داده به کار گرفته شده‌اند.

جدول ۱ شامل جزئیات این مجموعه داده‌ها است.

## ۲.۵ معیارهای ارزیابی

برای ارزیابی عملکرد الگوریتم پیشنهادی و نیز الگوریتم مورد استفاده در مقایسات، زمان اجرای الگوریتم و ۲ معیار دسته‌بندی چندبرچسبی مورد استفاده قرار گرفته‌اند. این دو معیار دقت دسته‌بندی و زیان همینگ [۱۴] هستند که به صورت زیر تعریف می‌شوند.

فرض کنیم  $T = \{(x_i, Y_i), i = 1, \dots, p\}$  مجموعه آزمون،  $Y_i \subseteq L$  مجموعه برچسب واقعی و  $Z_i \subseteq L$  مجموعه برچسب‌های پیش‌بینی شده برای نمونه  $x_i$  باشند. هم‌چنین  $f(x, y)$  را امتیاز در نظر گرفته شده به برچسب  $y$  برای نمونه  $x$  در نظر بگیریم. بنابراین داریم:

تعریف ۱.۵ (زیان همینگ). این معیار تعداد دفعاتی که یک جفت نمونه-برچسب به درستی دسته‌بندی نشده است را محاسبه می‌کند:

$$HammingLoss(T) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \Delta Z_i|}{|L|} \quad (18)$$

که  $\Delta$  تقارن متقارن بین دو مجموعه و  $|L|$  برابر تعداد برچسب‌ها است.

تعریف ۲.۵ (دقت). این معیار درصد برچسب‌هایی که به درستی پیش‌بینی شده‌اند را

محاسبه می‌کند:

$$Accuracy(T) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (19)$$

### ۳.۵ نتایج آزمایش‌ها

همان‌گونه که در بخش قبل اشاره شد، روش پیشنهادی با ۵ روش انتخاب ویژگی چندبرچسبی و بر اساس ۵ مجموعه داده دنیای واقعی ارزیابی شده است. پارامترهای این روش‌ها بر اساس مقادیر پیش‌بینی شده در مقالات آنها تنظیم شده‌اند. برای فرآیند یادگیری نیز، دسته بند  $ML - KNN$  [۲۶] که نسخه چندبرچسبی دسته‌بند معروف  $k$  -همسایه نزدیک‌تر است، در نظر گرفته شده است. در این آزمایش‌ها تعداد همسایه‌ها برابر ۱۰ در نظر گرفته شده است. در هر آزمایش ۶۰ درصد از داده‌ها به صورت تصادفی به عنوان داده آموزشی و ۴۰ درصد باقی‌مانده آنها به عنوان داده آزمون در نظر گرفته شده‌اند. لازم به ذکر است این فرآیند برای هر روش ۲۰ مرتبه تکرار شده است و میانگین این ۲۰ تکرار به عنوان نتیجه نهایی گزارش شده است. نتایج مقایسه روش پیشنهادی با سایر الگوریتم‌های مشابه از نظر معیار دقت در شکل ۱، از نظر معیار زیان همینگ در شکل ۲ و از نظر زمان اجرا بر اساس اجرای هر الگوریتم بر حسب ثانیه در جدول ۲ قرار گرفته‌اند. لازم به ذکر است کلیه آزمایش‌ها بر روی یک کامپیوتر با سیستم عامل ویندوز سرور<sup>۳۴</sup> ۲۰۱۲ با ۶۴ گیگابایت حافظه رم و ۱۶ پردازنده  $IntelXeon\ gold\ 6254$  صورت گرفته است. همچنین تمام الگوریتم‌ها با زبان متلب<sup>۳۵</sup> و در نرم‌افزار  $MatlabR2020b$  اجرا شده‌اند. در هر شکل، پنج تصویر قرار دارد که نشان‌دهنده نتایج بر روی یک مجموعه داده است و نام مجموعه داده در بالای هر تصویر قرار گرفته است.

در این مقاله، روش پیشنهادی با ۵ روش قدرتمند در انتخاب ویژگی چندبرچسبی مقایسه شده‌اند. لازم به ذکر است در ۲ روش  $MFS - MCDM$  و  $PPT - ReliefF$  تنها همبستگی بین ویژگی‌ها و برچسب‌ها در نظر گرفته شده است و از افزونگی بین ویژگی‌ها صرف‌نظر شده است. اما در ۳ روش دیگر، علاوه بر همبستگی بین ویژگی‌ها و برچسب‌ها، شباهت بین ویژگی‌ها نیز در نظر گرفته شده است. با این حال در این روش‌ها

<sup>34</sup>Windows server

<sup>35</sup>Matlab



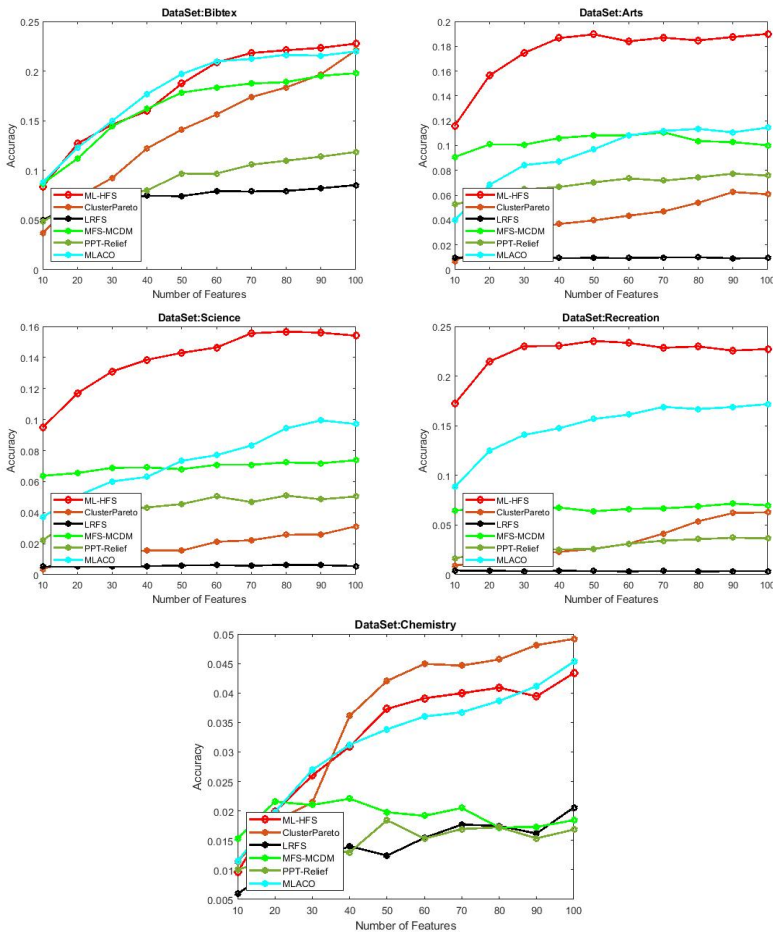
نیز همبستگی با برجسبها و شباهت بین ویژگیها تنها بر اساس يك معيار اندازه‌گیری شده است. با توجه به این که هر معيار، رفتار خاصی از ویژگیها را در نظر می‌گیرد و تجربه نشان داده است که ترکیبی از روش‌های مختلف غالباً از استفاده از يك معيار ثابت موثرتر است. بنابراین در روش پیشنهادی ما برای محاسبه هر کدام از همبستگی با برجسبها و شباهت ویژگیها، از سه معيار متفاوت استفاده کرده‌ایم. مقدار بدست آمده توسط هر روش به عنوان مجموعه‌های فازی مردد برای ویژگیها در نظر گرفته می‌شود. این مسئله در هیچ يك از روش‌های مقایسه شده صورت نگرفته است. بر اساس این استراتژی، روش پیشنهادی ما در دسته‌بندی ۲ درصد نسبت به دو روش *MLACO* و *MFS - MCDM*، ۴ درصد در مقابل *PPT - ReliefF*، ۶ درصد در برابر *ClusterPareto* و ۲۲ درصد نسبه به روش *LRFS* بهبود داشته است. همچنین از نظر زمان اجرا نیز پس از الگوریتم *MFS - MCDM* سریع‌ترین روش است و از نظر زمان اجرا نیز اختلاف ناچیزی با این روش دارد.

جدول ۲: زمان اجرای الگوریتمها

Science	Recreation	Chemistry	Bibtex	Arts	روش/مجموعه داده
۱/۱۰	۰/۸۹	۰/۸۶	۵۹۵	۰/۶۴	ML-HFS
۴/۸۴	۲/۸۸	۱۱/۳۰	۴۷/۰۳	۱/۹۱	ClusterPareto
۵/۹۹	۴/۰۸	۱۶/۴۱	۵۵/۱۹	۲/۸۸	LRFS
۰/۱۷	۰/۱۲	۰/۱۲	۱/۰۲۷	۰/۵۷	MFS-MCDM
۲۶/۲۴	۲۲/۰۸	۱۵/۱۱	۷۳/۹۴	۱۴/۸۱	PPT-ReliefF
۵/۴۴	۴	۳/۱۷	۴۶/۶۷	۱/۵۲	MLACO

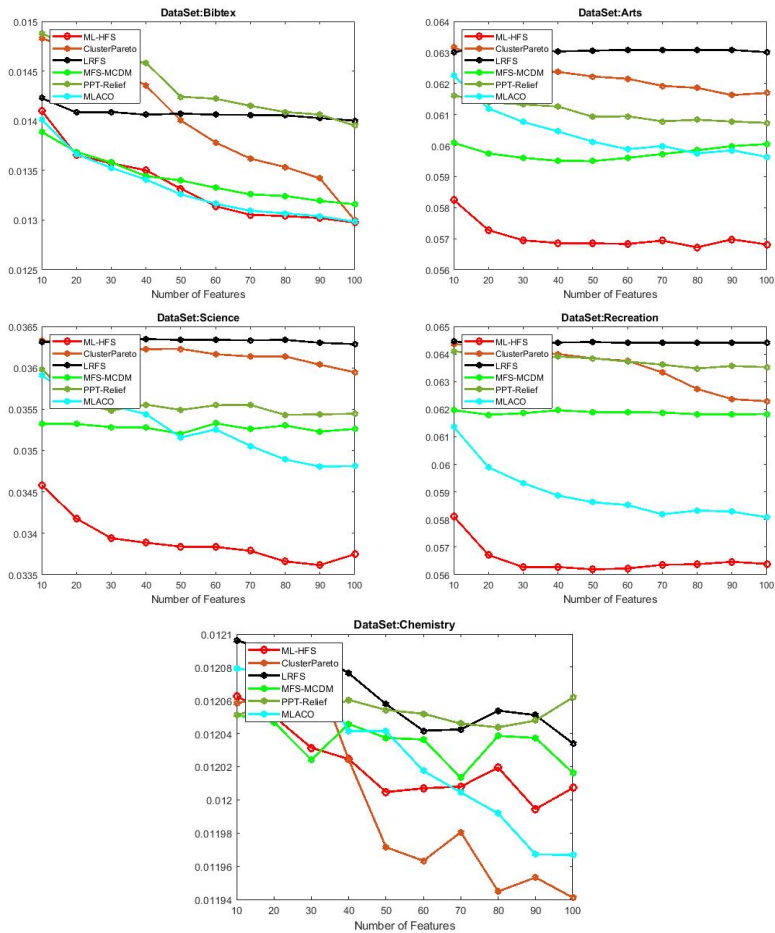
## ۶ نتیجه‌گیری و کارهای آینده

در این مقاله، ما يك الگوریتم جدید برای انتخاب ویژگی چندبرجسبی با استفاده از مجموعه‌های فازی ارائه کردیم. در این روش ما به دنبال حداکثر کردن همبستگی بین ویژگیها و برجسبهای کلاس و همچنین حداقل کردن افزونگی بین ویژگیها بودیم. از طرفی با توجه به این که هر معيار انتخاب ویژگی فقط جنبه خاصی از ویژگی را در نظر می‌گیرد. از این رو از يك رویکرد شورایی برای محاسبه میزان همبستگی و افزونگی ویژگیها استفاده کرده‌ایم. به این صورت که برای هر رویکرد، سه معيار انتخاب ویژگی بر



شکل ۱: نتایج بر اساس معیار دقت

اساس مفهوم انرژی اطلاعات در مجموعه‌های فازی مردد با هم ترکیب می‌شوند و در پایان بر اساس یک معیار شایستگی مقادیر بدست آمده برای همبستگی و افزونگی ویژگی‌ها با هم ترکیب می‌شوند و ویژگی‌ها بر اساس آن رتبه‌دهی می‌شوند. در آینده ما به دنبال این هستیم که از سایر عملگرهای فازی مردد در مسائل انتخاب ویژگی استفاده کنیم. همچنین با استفاده از مجموعه‌های فازی، رویکردهای مبتنی بر انتخاب ویژگی شورایی را ارائه کنیم. همچنین با توجه به رشد روزافزون داده‌های چندبرچسبی، هم‌چنان به دنبال ارائه روش‌های موثری برای این داده‌ها هستیم.



شکل ۲: نتایج بر اساس معیار زیان همینگ

## مراجع

- [۱] نصرتی ناهوک، حسن، افتخاری، مهدی. (۱۳۹۲) یک روش جدید برای انتخاب ویژگی مبتنی بر منطق فازی. سیستم‌های هوشمند در مهندسی برق، دوره ۴، شماره (۱): صص. ۷۱ تا ۸۳.

[2] Atanassov, K.T. (1986) Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, **20**, 87–96.

[3] Bolón-Canedo, V. and Alonso-Betanzos, A. (2019) Ensembles for feature selec-

- tion: A review and future trends. *Information Fusion*, **52**, 1–12.
- [4] Cai, J., Luo, J., Wang, S. and Yang, S. (2018) Feature selection in machine learning: A new perspective. *Neurocomputing*, **300**, 70–79.
- [5] Ebrahimpour, M.K. and Eftekhari, M. (2017) Ensemble of feature selection methods: A hesitant fuzzy sets approach. *Applied Soft Computing Journal*, **50**, 300–312.
- [6] Gong, J.-W., Liu, H.-C., You, X.-Y. and Yin, L. (2021) An integrated multi-criteria decision making approach with linguistic hesitant fuzzy sets for E-learning website evaluation and selection. *Applied Soft Computing*, **102**, 107118.
- [7] Hall, M.A. (1999) Correlation-based feature selection for machine learning. *University of Waikato*.
- [8] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2021) A pareto-based ensemble of feature selection algorithms. *Expert Systems with Applications*, **180**, 115130.
- [9] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-Pour, H. (2021) An efficient Pareto-based feature selection algorithm for multi-label classification. *Information Sciences*, **581**, 428–447.
- [10] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-Pour, H. (2021) VMFS: A VIKOR-based multi-target feature selection. *Expert Systems with Applications*, 115224.
- [11] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2021) Ensemble of feature selection algorithms: a multi-criteria decision-making approach. *International Journal of Machine Learning and Cybernetics*, 1–21.

- [12] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2020) MFS-MCDM: Multi-label feature selection using multi-criteria decision making. *Knowledge-Based Systems*, **206**, 106365.
- [13] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2020) MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality. *Expert Systems with Applications*, **142**, 113024.
- [14] Hashemi, A., Dowlatshahi, M.B. and Nezamabadi-Pour, H. (2021) A bipartite matching-based feature selection for multi-label learning. *International Journal of Machine Learning and Cybernetics*, **12**, 459–475.
- [15] Joodaki, M., Dowlatshahi, M.B. and Joodaki, N.Z. (2021) An ensemble feature selection algorithm based on PageRank centrality and fuzzy logic. *Knowledge-Based Systems*, **233**, 107538.
- [16] Kashef, S. and Nezamabadi-pour, H. (2019) A label-specific multi-label feature selection algorithm based on the Pareto dominance concept. *Pattern Recognition*, **88**, 654–667.
- [17] Kashef, S., Nezamabadi-pour, H. and Nikpour, B. (2018) Multilabel feature selection: A comprehensive review and guiding experiments. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**, e1240.
- [18] Paniri, M., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2021) Ant-TD: Ant colony optimization plus temporal difference reinforcement learning for multi-label feature selection. *Swarm and Evolutionary Computation*, **64**, 100892.
- [19] Paniri, M., Dowlatshahi, M.B. and Nezamabadi-pour, H. (2020) MLACO: A multi-label feature selection algorithm based on ant colony optimization. *Knowledge-Based Systems*, **192**, 105285.

- [20] Pereira, R.B., Plastino, A., Zadrozny, B. and Merschmann, L.H.C. (2018) Categorizing feature selection methods for multi-label classification. *Artificial Intelligence Review*, **49**, 57–78.
- [21] Reyes, O., Morell, C. and Ventura, S. (2015) Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. *Neurocomputing*, **161**.
- [22] Rickard, J.T., Aisbett, J. and Gibbon, G. (2009) Fuzzy subsethood for fuzzy sets of type-2 and generalized type-n. *IEEE Transactions on Fuzzy Systems*, **17**, 50–60.
- [23] Torra, V. (2010) Hesitant fuzzy sets. *International Journal of Intelligent Systems*, **25**, 529–539.
- [24] Venkatesh, B. and Anuradha, J. (2019) A review of Feature Selection and its methods. *Cybernetics and Information Technologies*, **19**, 3–26.
- [25] Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338–353.
- [26] Zhang, M.L. and Zhou, Z.H. (2007) ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, **40**, 2038–2048.
- [27] Zhang, P., Liu, G. and Gao, W. (2019) Distinguishing two types of labels for multi-label feature selection. *Pattern Recognition*, **95**, 72–82.
- [28] Zhang, R., Nie, F., Li, X. and Wei, X. (2019) Feature selection with multi-view data: A survey. *Information Fusion*, **50**, 158–167.