

## ارزیابی قوانین انجمنی در داده کاوی

مهرداد مدهوشی<sup>۱</sup>، عباس نامدار علی آبادی<sup>۲</sup>

### چکیده

یکی از قدرتمندترین تکنیک‌ها برای یافتن الگو در داده‌ها، داده‌کاوی است. در این مقاله، روشی برای ارزیابی قوانین انجمنی از نقطه نظر تئوری اطلاعات پیشنهاد شده است که محاسبه آن را بر اساس مشخصه‌های شناخته شده قوانین انجمنی یعنی سه پارامتر، حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) امکان‌پذیر می‌سازد.

روش مورد نظر اجازه می‌دهد یک مدل خطی روی داده‌ها ایجاد کنیم که برای ارزیابی و فیلترینگ مفید است. در کنار آن تحلیل ارزش خود توصیفی قانون انجمنی روی سه پارامتر انجام شده است.

این مدل به ما اجازه می‌دهد قوانین انجمنی با ارزش‌های متفاوت از  $C_1$  و  $C_2$  و  $C_3$  را با هم مقایسه کنیم خود توصیفی یک قانون انجمنی می‌تواند برای فیلتر کردن و کشف قانون در پایگاه داده استفاده گردد. پیدا کردن قانون انجمنی با ارزش‌های پائین در  $C_1$  و  $C_2$  و  $C_3$  هزینه‌های اضافی در بردارد که به تحقیقات جداگانه‌ای نیازمند است.

**کلیدواژه:** قوانین انجمنی، ارزیابی، داده‌کاوی

## Evaluation of Association Rules in Data mining

Mehrdad Madhoshi, Abbas Nnamdar Aliabadi

### Abstract

In this paper a method for evaluating *association rule* from the view point of information theory has been suggested which allows us to calculate a generalized characteristic of *association rule* with three parameters: support ( $c^1$ ), improvement ( $c^2$ ) and confidence ( $c^3$ ). One of the most powerful techniques in finding out the patterns with respect to data mart is data mining.

The proposed method allows setting a linear order on the *association rule* which is useful for evaluating and filtering obtained the *rules*. Besides we have carried out analysis the self-descriptiveness on the three parameters. This paper allows us to compare *association rules* with different values of  $c^1$ ,  $c^2$ ,  $c^3$ . The self-descriptiveness of a rule can be used to filter the *association rules* discovered in a database. Of course finding rules with bad values of  $c^1$ ,  $c^2$ ,  $c^3$  is associated with additional expenses and problems associated with the necessity of finding such parameters require separate research for particular tasks.

**Key words:** Evaluation, Association Rules, Data mining

<sup>۱</sup>. دانشیار گروه مدیریت بازرگانی دانشگاه مازندران (Email: [Madhoshi@umz.ac.ir](mailto:Madhoshi@umz.ac.ir))

<sup>۲</sup>. دانشجوی کارشناسی ارشد مدیریت اجرایی دانشگاه مازندران (Email: [Namdar1980@gmail.com](mailto:Namdar1980@gmail.com))

## ۱- مقدمه

استخراج قوانین انجمنی نوعی عملیات داده کاوی است که به جستجو برای یافتن ارتباط بین خصوصیات در پایگاه داده می‌پردازد، به عبارت دیگر مطالعه خصوصیات می‌باشد که با یکدیگر همراه هستند [۳، ۴]. قوانین انجمنی به شکل اگر آنگاه هستند، تعریفی که برای قوانین انجمنی آمده چنین است [۲]:

اگر  $L = I_1, \dots, I_m$  مجموعه‌ای از خصوصیات شی باشد و  $T$  مجموعه‌ای از رکوردها در پایگاه داده باشد، هر رکورد  $t$  با یک بردار دو گانه (دودویی)  $t_{[K]} = 1$  بیان می‌گردد اگر رکورد  $t$  در برگرفته‌ی خصوصیات  $I_k$  ( $K=1, \dots, m$ ) باشد و  $t_{[K]} = 0$  اگر  $t$  شامل خصوصیات  $I_k$  نباشد. چنانچه  $X$  زیر مجموعه‌ای از خصوصیات مجموعه‌ی  $L$  به گونه‌ای که  $X \subseteq L$  می‌گوئیم که رکورد  $t$  واجد خصوصیت  $X$  است در صورتی که

$$\forall I_k \in X, t_{[K]} = 1 \text{ در صورتی که } X \rightarrow Y$$

$$X \cap Y = \emptyset, Y \subseteq L, X \subseteq L$$

پارامترهای مختلفی در طی سال‌های گذشته برای ارزیابی قوانین انجمنی پیشنهاد گردیده‌اند که از آن جمله حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ )، را می‌توان نام برد [۵].

منطقی است که یک شاخص گسترش یابد تا قابلیت بررسی چندین پارامتر را با هم در آن واحد داشته باشد برای مثال چگونه می‌توان در قانون انجمنی را که یکی از آنها از سطح بهبود و حمایت بالاتری برخوردار است با دیگری که سطح اعتماد بالاتری دارد مقایسه کرد.

کدام قانون انجمنی بهتر است؟ و چه قدر بهتر است؟

بسیار جالب خواهد بود که چنین شاخصه‌ای را برای تحلیل قوانین انجمنی پیدا کرد، چنین شاخصی با کمک ۳ پارامتر فوق‌الذکر محاسبه شده و هنگامی که قوانین انجمنی قویتر هستند افزایش خواهد یافت و در صورتی که قوانین انجمنی ضعیفتر باشند، کاهش می‌یابد.

## ۲- خصوصیات قوانین انجمنی از دیدگاه احتمالات

حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) به عنوان پارامترهای قابل قبول برای قوانین انجمنی شناخته می‌شود [۵].

از آن جایی که قوانین انجمنی طبیعتاً احتمالی می‌باشد [۱]، بنابراین می‌توان آنها را از دیدگاه احتمالات بررسی کرد. ارزش احتمال سطح حمایت ( $X \rightarrow Y$ ) به این صورت تعریف می‌شود:

نسبت تعداد رکوردها در پایگاه داده که در ( $X \rightarrow Y$ ) می‌گنجند به کل رکوردهای پایگاه داده.

همچنین می‌توان آن را به عنوان تعداد رکوردهایی که در قانون انجمنی مورد نظر صدق می‌کنند تعریف کرد. از دیدگاه احتمالات سطح حمایت منعکس کننده احتمال این حقیقت است که یک شیء در پایگاه داده دارای دو خصوصیت  $X$  و  $Y$  می‌باشد.

$$P(X \& Y) = P(XY)$$

و احتمال سطح اعتماد به صورت نسبت تعداد رکوردها که در معادله زیر صدق می‌کنند تعریف می‌شود. از دیدگاه احتمالات سطح اعتماد منعکس کننده احتمال شرطی زیر می‌باشد.  $C_3(X \rightarrow Y)$

$$P_X(Y) = \frac{P(XY)}{P(X)}$$

اگر ( $X \rightarrow Y$ ) یک قانون انجمنی باشد، در این صورت شرایط زیر برقرار است.

$$C_1(X \rightarrow Y) = P(XY) \geq \min C_1;$$

$$\frac{C1(X \rightarrow Y)}{C1(X)} = P_X(Y) \geq \min C_2,$$

در این معادله کمترین مقدار سطح حمایت و کمترین مقدار سطح اعتماد نشان داده شده‌اند، سطح بهبود به صورت زیر نشان داده می‌شود.

$$C_2(X \rightarrow Y) = \frac{P_X(Y)}{P(Y)} = \frac{C2(X \rightarrow Y)}{C1(Y)}$$

بنابراین سطح بهبود نشان دهنده‌ی افزایش در احتمال حادثه‌ی  $Y$  تحت شرط وقوع حادثه‌ی  $X$  است که در این حالت احتمال  $Y$  غیر مشروط است.

در ادامه نشان داده می‌شود که ۳ پارامتر فوق‌الذکر این اجازه را به ما می‌دهند که آنها را در یک خصوصیت دودویی بیان کنیم.

اگر هر دو مقدار عددی فقط ارزش‌های ۰ و ۱ به خود بگیرند، ما می‌توانیم احتمالات  $P_{00}, P_{01}, P_{10}, P_{11}$  را بنویسیم که تمامی ترکیبات ممکن از این دو مقدار را به ما می‌دهد که می‌توان آنها را با در نظر گرفتن حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) بیان کرد.

ابتدا می خواهیم احتمالات موجود در رابطه با این ۳ پارامتر را در ارزش  $(X \rightarrow Y)$  نشان دهیم .

$$C_{\setminus} (X \rightarrow Y) = P(XY) = P_{11};$$

$$C_{\setminus} (X \rightarrow Y) = \frac{C1(X \rightarrow Y)}{C1(X)} = \frac{P(XY)}{P(X)} = \frac{P_{11}}{P_{11} + P_{10}}.$$

$$C_{\setminus} (X \rightarrow Y) = \frac{C1(X \rightarrow Y)}{C1(X)C1(Y)} = \frac{P(XY)}{P(X)P(Y)} = \frac{P_{11}}{(P_{11} + P_{10})(P_{11} + P_{01})}$$

با در نظر گرفتن این عبارات ارزش های احتمالی زیر به دست می آید .

$$P_{11} = C_{\setminus} (X \rightarrow Y); \quad (1)$$

$$P_{1.} = \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - C_{\setminus} (X \rightarrow Y) = C_{\setminus} (X \rightarrow \bar{Y}); \quad (2)$$

$$P_{.1} = \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)} - C_{\setminus} (X \rightarrow Y) = (\bar{X} \rightarrow Y); \quad (3)$$

$$P_{..} = 1 - P_{11} - P_{1.} - P_{.1} =$$

(4)

$$= 1 + C_{\setminus} (X \rightarrow Y) - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)} = C_{\setminus} (\bar{X} \rightarrow \bar{Y}).$$

بنابراین برای محاسبه  $P_{11}$  ,  $P_{1.}$  ,  $P_{.1}$  ,  $P_{..}$  ما از تمامی ۳ پارامتر موجود استفاده کردیم . بدیهی است که اگر هر یک از این پارامتر ها موجود نباشد ارزش های احتمالی قابل دستیابی نیستند .

شایان ذکر است که این ۳ پارامتر را نمی توان به صورت تصادفی تعیین کرد . این ۳ پارامتر باید در شرایط زیر صدق کنند.

$$0 \leq C_{\setminus} (X \rightarrow Y) \leq 1. \quad (5)$$

$$C_{\setminus} (X \rightarrow Y) \leq C_{\setminus} (X \rightarrow Y) \leq 1. \quad (6)$$

$$\frac{(C2(X \rightarrow Y))^2}{C1(X \rightarrow Y).C2(X \rightarrow Y) + C2(X \rightarrow Y) - C1(X \rightarrow Y)} \leq C_{\setminus} (X \rightarrow Y) \leq \frac{C2(X \rightarrow Y)}{C1(X \rightarrow Y)}. \quad (7)$$

نا مساوی های بالا به راحتی قابل اثبات است .

### ۳- ارزیابی قوانین انجمنی با توجه به تئوری اطلاعات

اکنون روشی برای ارزیابی قوانین انجمنی با توجه به تئوری اطلاعات ارائه می شود . با در نظر گرفتن اینکه قوانین انجمنی  $(X \rightarrow Y)$  می تواند به عنوان اطلاعاتی به روی پیشامد  $Y$  منظور شود که از دریافت پیامی درباره ی پیشامد  $Y$  منتج می شود . این اطلاعات « از پیشامد به پیشامد » به صورت این فرمول تعریف می شود .

$$I_{X \leftrightarrow Y} = \log_{\setminus} \frac{P(XY)}{P(X)P(Y)}$$

به عبارت دیگر

$$\log_{\setminus} (C_{\setminus} (X \rightarrow Y)).$$

اطلاعات دو طرفه به صورت این فرمول ارائه می گردد .

$$I_{X \leftrightarrow Y} = \sum_{i=1}^n \sum_{j=1}^m P_{ij} \log_{\setminus} \frac{P_{ij}}{P_i P_j},$$

که در این فرمول  $P_{ij}$  احتمال این واقعیت است که  $X$  در حالت  $x_i$  و  $Y$  در حالت  $y_j$  است .

$$P_{ij} = P((X \sim x_i)(Y \sim y_j))$$

$P_i = p(x \sim x_i)$  احتمال واقعی است که  $x$  در حالت  $x_i$  است .

$P_j = p(y \sim y_j)$  احتمال واقعی است که  $y$  در حالت  $y_j$  است .

در مسئله ی مورد نظر ما  $p_{ij} \in \{P_{..}, P_{.1}, P_{1.}, P_{11}\}$

با در نظر گرفتن مطالب بالا می توان فرمولی برای اطلاعات دو طرفه ی یک قانون انجمنی ارائه دهیم که نام این فرمول خود توصیفی قانون انجمنی می باشد. (۸)

$$I_{X \leftrightarrow Y} = C_1(X \rightarrow Y) \cdot \log_2(C_2(X \rightarrow Y)) \\ + C_1(X \rightarrow \bar{Y}) \cdot \log_2(C_2(X \rightarrow \bar{Y})) + C_1(\bar{X} \rightarrow Y) \cdot \log_2(C_2(\bar{X} \rightarrow Y)) \\ + C_1(\bar{X} \rightarrow \bar{Y}) \cdot \log_2(C_2(\bar{X} \rightarrow \bar{Y})).$$

همان گونه که قبلاً بحث شد ۳ پارامتر حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) تمام احتمالات را توصیف می کند .

بنابراین تمامی پارامترها در فرمول ۸ می توانند با حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) بیان گردد .

حمایت از فرمول های ۴-۱ به دست می آید . اعتماد نیز به همین ترتیب

در فرمول های زیر خصوصیات قانون انجمنی فوق اعمال شده :

$$C_2(\bar{X} \rightarrow Y) = \frac{C1(\bar{X} \rightarrow Y)}{C1(X)C1(Y)} = \frac{P_{01}}{(P_{01} + P_{00})(P_{01} + P_{11})}; \quad (9)$$

$$C3(X \rightarrow \bar{Y}) = \frac{C1(X \rightarrow \bar{Y})}{C1(X)C1(\bar{Y})} = \frac{P_{10}}{(P_{10} + P_{11})(P_{10} + P_{00})}; \quad (10)$$

$$C3(\bar{X} \rightarrow \bar{Y}) = \frac{C1(\bar{X} \rightarrow \bar{Y})}{C1(\bar{X})C1(\bar{Y})} = \frac{P_{00}}{(P_{00} + P_{01})(P_{00} + P_{01})}. \quad (11)$$

با جایگزینی عبارات  $P_{..}, P_{.1}, P_{1.}, P_{11}$  از فرمول ۴-۱ در فرمول های ۱۱-۹ و انجام تبدیلات جبری به فرمول های زیر دست می یابیم .

$$C3(\bar{X} \rightarrow Y) = \frac{C2(X \rightarrow Y) - C1(X \rightarrow Y) \cdot C3(X \rightarrow Y)}{C2(X \rightarrow Y) - C1(X \rightarrow Y)}; \quad (12)$$

$$C3(X \rightarrow \bar{Y}) = \frac{C3(X \rightarrow Y)(1 - C2(X \rightarrow Y))}{C3(X \rightarrow Y) - C2(X \rightarrow Y)}; \quad (13)$$

$$C3(\bar{X} \rightarrow \bar{Y}) = \frac{1 + C1(X \rightarrow Y) - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)}}{\left[1 - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)}\right] \left[1 - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)}\right]}. \quad (14)$$

حال پس از انجام این عملیات به فرمولی برای اطلاعات دو طرفه دست پیدا می کنیم که خصوصیات قانون انجمنی (اعتماد ، حمایت و بهبود) در آن لحاظ شده است .

$$I_{X \leftrightarrow Y} = C1(X \rightarrow Y) \cdot \log_2(C3(X \rightarrow Y)) \\ + \left[ \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - C1(X \rightarrow Y) \right] \cdot \log_2 \left[ \frac{C3(X \rightarrow Y)(1 - C2(X \rightarrow Y))}{C3(X \rightarrow Y) - C2(X \rightarrow Y)} \right] \\ + \left[ \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)} - C1(X \rightarrow Y) \right] \\ \times \log_2 \left[ \frac{C2(X \rightarrow Y) - C1(X \rightarrow Y) \cdot C3(X \rightarrow Y)}{C2(X \rightarrow Y) - C1(X \rightarrow Y)} \right]$$

$$+ \left[ 1 + C1(X \rightarrow Y) - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)} \right] \\ \times \log_2 \left[ \frac{1 + C1(X \rightarrow Y) - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)}}{\left[ 1 - \frac{C1(X \rightarrow Y)}{C2(X \rightarrow Y)} \right] \left[ 1 - \frac{C2(X \rightarrow Y)}{C3(X \rightarrow Y)} \right]} \right]$$

(۱۵)

شایان ذکر است که عبارات در لگاریتم باید بزرگتر یا مساوی صفر باشند. چنانچه به صفر رسیدیم از حد استفاده می کنیم تا محاسبات لازم را انجام دهیم.

قابل ذکر است که خود توصیفی قانون انجمنی برای هر ارزش احتمالی قابل محاسبه است و باعث می شود بتوان از این ارزش به دست آمده به عنوان آستانه ای برای فیلتر کردن استفاده شود.

اگر  $X$  و  $Y$  مستقل باشند و رابطه ای بین آنها وجود نداشته باشد آنگاه خود توصیفی صفر خواهد بود. بیشترین مقدار خود توصیفی ۱ خواهد بود. در دنیای واقعی دستیابی به چنین پارامترهایی به سختی ممکن است. مسئله ی تعریف آستانه های عددی برای قانون انجمنی در پایگاه های داده نیاز به تحقیقات بیشتری دارد.

## ۴- وابستگی خود توصیفی به پارامتر های قانون انجمنی

حال چگونگی تاثیر حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) را بر خود توصیفی بررسی می کنیم واضح است که هر چه میزان ۳ پارامتر فوق بیشتر باشد ما خود توصیفی بیشتر و بهتری داریم. به منظور تحلیل اینکه چگونه خود توصیفی بستگی به ۳ پارامتر دارد ابتدا به فرمول ۱۵ بر می گردیم که یک تابع متفاوت با ۳ پارامتر تعریف می کند.

هنگامی که ارزش حمایت ( $C_1$ ) و بهبود ( $C_2$ ) ثابت هستند در این حالت اعتماد ( $C_3$ ) مطابق با نا مساوی ۶ تغییر می کند. مشخصاً تابع تعریف شده دارای یک مقدار مینیمم است به منظور پیدا کردن ارزش اعتماد ( $C_3$ ) برای ما که مطابق با Min تابع باشد باید از آن مشتق گرفت.

$$\frac{\partial I_{X \leftrightarrow Y}}{\partial C2} = \frac{C1}{C2^2} \\ \times \log_2 \left[ \frac{C2.C3 + C1.C2.C3 - C1.C3 - Conf^2}{C2.C3 + C1.C2.C3 - C1.C3 - C2^2.C3} \right] \\ + \frac{1}{C3} \cdot \log_2 \left[ \frac{C2.C3 + C1.C2.C3 - C1.C3^2 - C2^2}{C2.C3 + C1.C2.C3 - C1.C3 - C2^2} \right]$$

مشتق مورد نظر هنگامی برابر با صفر است که  $C_3 = \sqrt{C_1.C_2}$

به عنوان مثال اگر  $C_1 = 0.1$  و  $C_2 = 2$  تابع می نیمم خواهد بود اگر  $C_3 = 0.447$  بنابراین یک قانون با سطح اعتماد بالاتر لزوماً دارای ارزش خود توصیفی بالاتری نمی باشد. یعنی اینکه قانون الزاماً اطلاعات بیشتری از  $X$  به  $Y$  گسیل نمی دارد.

برای مثال هنگامی که ( $C_1 = 0.3$ ،  $C_2 = 2$ ) یک قانون با سطح اعتماد ( $C_3 = 0.6$ ) دارای خود توصیفی برابر  $I_{X \leftrightarrow Y} = 0.6$  (اطلاعات بیشتری گسیل می دارد نسبت به یک قانون که سطح اعتماد آن برابر  $0.78$  و خود توصیفی آن  $I_{X \leftrightarrow Y} = 0.3$ ) است.

اگر چه سطح اعتماد دومین قانون  $1/3$  برابر، بزرگتر است، اما با در نظر گرفتن تاثیر حمایت بر خود توصیفی، خود توصیفی کمتری دارد. اکنون نشان می دهیم که چگونه حمایت و بهبود خود توصیفی یک قانون را تحت تاثیر قرار می دهد. با در نظر گرفتن تاثیر حمایت بر خود توصیفی یک قانون و مشتق تابع اطلاعات دوطرفه، و اعتماد و بهبود ثابت به فرمول زیر دست می یابیم:

$$\frac{\partial I_{X \leftrightarrow Y}}{\partial C1} = \frac{1}{C2} \times \log_2 \left[ \frac{C3(1-C2)(C2-C1)}{C2.C3 + C1.C2.C3 - C1.C3 - C2^2} \right] - \log_2 \left[ \frac{(1-C2)(C2-C1.C3)}{C2.C1 + C1.C2.C3 - C1.C3 - C2^2} \right]$$

$$C_1 > 1, \frac{\partial I_{X \leftrightarrow Y}}{\partial C1} > 0.$$

حال در فرمول زیر، چگونگی تاثیر بهبود بر خود توصیفی نشان داده می شود:

$$\frac{\partial I_{X \leftrightarrow Y}}{\partial C3} = \frac{C2}{C3^2} \times \log_2 \left[ \frac{C2.C3 + C1.C2.C3 - C1.C3 - C2^2}{C3 - C2} \right] - \frac{C2}{C3} . \log_2 (C2 - C1.C3).$$

این مشتق هنگامی که  $C_1=1$  به صفر می رسد که بیانگر این مطلب است که خصوصیات غیر وابسته هستند در حالتی که هیچ اطلاعاتی از  $X$  به  $Y$  گسیل نمی شود. بهبود در حیطه ی حد های تعریف شده توسط نا مساوی  $Y$  تغییر می یابد . شایان ذکر است که حمایت و بهبود، پارامتر های متقارن و یک جهتی هستند، و این مطلب در مورد خود توصیفی یک قانون نیز صحیح است و از نقطه نظر تئوری اطلاعات این مسئله امری طبیعی است . چرا که اطلاعات دو طرفه از  $X$  به  $Y$  و از  $Y$  به  $X$  همواره یکسان است. این مطلب این حقیقت را روشن می کند که چنانچه حمایت و بهبود رشد کند، خود توصیفی نیز رشد می کند اما اعتماد یک پارامتر متقارن نیست. اعتماد خود تعیین کننده ی جهت خود توصیفی است.

## ۵- نتیجه گیری

روش پیشنهادی برای ارزیابی قوانین انجمنی از نقطه نظر تئوری اطلاعات بر اساس محاسبه ی یک پارامتر جمعی از اطلاعات دو طرفه می باشد که به ما اجازه می دهد قوانین انجمنی را با ارزش های مختلف حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) که از یک پایگاه داده به دست می آید مقایسه کنیم . استفاده از چنین پارامتری لازمی رد کردن پارامترهای متداول نیست بلکه مکمل آنها بوده و اجازه می دهد این ارزش ها مد نظر قرار داده شود. خود توصیفی یک قانون می تواند برای فیلتر کردن قوانین انجمنی در یک پایگاه داده مورد استفاده قرار گیرد . تحلیل وابستگی خود توصیفی قوانین انجمنی بر خصوصیات حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) نشان می دهد که این وابستگی روشن و واضح نمی باشد . یعنی اینکه هر چه سطوح حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) بالاتر باشد، سطح خود توصیفی لزوماً بالاتر نیست . بنابراین قوانینی با ارزش های کم در حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) می تواند از نظر انتقال اطلاعات جالب باشند . البته یافتن قوانینی با ارزش های کم در حمایت ( $C_1$ )، بهبود ( $C_2$ ) و اعتماد ( $C_3$ ) در برگزیده های هزینه های اضافی (زمان و حافظه) بوده و نیاز به تحقیق بیشتری دارد . لازم به ذکر است که روش ارائه شده برای اطلاعات در یک پایگاه داده که به صورت بولین و باینری ذخیره شده اند به کار می رود.

## ۶- منابع

[۱] Agraval R, Imielinski T., Swami A. (۱۹۹۳). *Mining association rules between sets of items in large databases: Proc. of the ACM SIGMOD Conference*. Washington DC, USA, May ۱۹۹۳.-P.۲۰۷-۲۱۶.

[٢] Dunham M.H.(٢٠٠٢) *Data Mining , Introductory and advanced topics* , prentice hall.

[٣] Han. J,Kamber.M.(٢٠٠٦) *Data Mining concepts and techniques* , Morgan Kaufmann publishers

[٤] Klossgen.W.and Zytkow.J.M.(٢٠٠٢) *Handbook of Data Mining and Knowledge Discovery*,Oxford university press.

[٥] Sitnikov D, Titova E. & Ryabov O. (٢٠٠٦). *A method for association rule quality evaluation based on information theory: Data mining VII*, ٢٥-٣٤