یک شاخص پیچیدگی مبتنی بر انتخاب و تودرتویی

چکیده :

بسیاری از مفاهیم در خصوص کمی سازی و سنجش پیچیدگی برنامه در طی چند سال اخیر، توسعه یافته است. یکی از پذیرفته شده ترین و آسان ترین شاخص های پیچیدگی، موسوم به عدد سیکلوماتیک مک کیب، در مطالعات مختلف مورد بحث قرار گرفته و بهبود یافته است. عدد سیکلوماتیک تنها ساختار تصمیم یک برنامه را در نظر می گیرد. از این روی، این مقاله یک روش جدید را برای محاسبه پیچیدگی برنامه موسوم به مفهوم postdomination (پسا غالبیت) پیشنهاد می کند. این روش درجه تودرتویی یک برنامه را در نظر می گیرد. با ترکیب این روش و عدد سیکلوماتیک، یک شاخص پیچیدگی جدید تعریف خواهد شد.

کلمات کلیدی : عدد سیکلوماتیک، پیچیدگی برنامه، درجه تو در تویی، غالبیت پیشرو

1- مقدمه

از میان بسیاری از موارد پیشنهادی در خصوص کمی سازی پیچیدگی برنامه، مفهوم مک کیب(MCCA7)، ظاهرا یکی از پذیرفته شده ترین و آسان ترین شاخص های پیچیدگی است. ایشان پیشنهاد می کند که پیچیدگی ارتباط تنگاتنگی با اندازه برنامه ندارد، بلکه ارتباط نزدیکی با تعداد مسیر های پایه از طریق یک گراف کنترل برنامه دارد. مفهوم مک کیب از یک مدل گراف جهت دار برنامه ها و مبانی نظریه گراف برای محاسبه شاخص پیچیدگی بهره می برد. برای یک برنامه معین، ابتدا یک گراف کنترل(گراف جهت دار) ترسیم می شود. یک گره در گراف با چندین عبارت متناظر است و یک قوس یا یال با جریان کنترل محتمل در میان گره های مختلف متناظر است. برای این گراف G، عدد سیکلوکلیماتیک v(G) به صورت زیر تعریف می شود

$$v(G) = e - n + 2p,$$

که C تعداد یال ها،n تعداد گره ها و p تعداد مولفه های همبندی است.

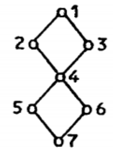
از این روی، پیچیدگی یک برنامه که با v(G) اندازه گیری می شود، فرضا تنها یک عامل از ساختار تصمیم برنامه است. با این حال، چندین ناهنجاری یافته شده است که در آن پیچیدگی بالاتر برای یک برنامه با پیچیدگی کم تر نسبت به یک برنامه با پیچیدگی بالاتر محاسبه می شود.

میرز برای اصلاح این ناهنجاری ها، محاسبه v(G) را به عنوان یک بازه پیچیدگی پیشنهاد می کند(MYER77). کران پایین بازه، به صورت تعداد عبارت های تصمیم به علاوه یک تعریف می شود(مثال ها برای عبارت های تصمیم شامل IF، DO WHILE و عبارت های DO تکراری) و کران بالاتر، تعداد شرایط فردی به علاوه یک است. هانسن یک شاخصی را ارایه کرده است که ترکیبی از عدد سیکلوماتیک و یک شمارش عملیات(HANS78) می باشد. از سوی دیگر، شاخص مک کیب ، پیچیدگی تودرتویی را در نظر نمی گیرد. چن، پیچیدگی یک برنامه را با شاخص MIN توصیف می کند(حداکثر عدد اشتراک،CHEN78).هاریسون و همکاران پیچیدگی برنامه ها را با بزرگ ترین کران پایین(GLB) گره انتخاب(HARR81) اندازه گیری می کنند. با این حال، مفهوم GLB گره انتخاب هنوز روشن نشده است.

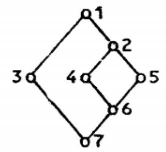
این مقاله به معرفی یک روش جدید برای محاسبه پیچیدگی، یعنی مفهوم postdomination که درجه تو درتویی را در نظر می گیرد می پردازد. با کمک این مفهوم، ما عدد سیکلوماتیک مک کیب را اصلاح کرده و یک شاخص پیچیدگی جدید را تعریف می کنیم.

از این روی، ما پنج برنامه زیر و گراف های کنترل آن ها را در نظر می گیریم:

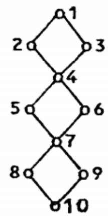
```
(a)     IF P1 THEN S2;
              ELSE S3;
        IF P4 THEN S5;
              ELSE S6;
        S7;
```



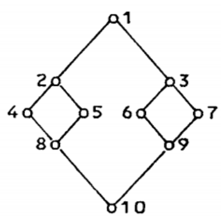
```
(b)     IF P1 THEN
                   IF P2 THEN S4;
                         ELSE S5;
                   S6;
              ELSE S3;
        S7;
```



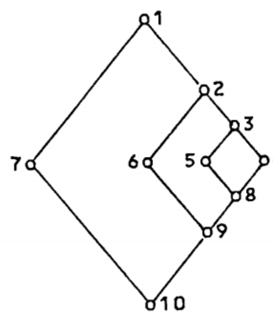
```
(c)     IF P1 THEN S2;
              ELSE S3;
        IF P4 THEN S5;
              ELSE S6;
        IF P7 THEN S8;
              ELSE S9;
        S10;
```



```
(d)     IF P1 THEN
                   IF P2 THEN S4;
                         ELSE S5;
                   S8;
              ELSE
                   IF P3 THEN S6;
                         ELSE S7;
                   S9;
        S10;
```



```
(e)     IF P1 THEN
                   IF P2 THEN
                              IF P3 THEN S4;
                                    ELSE S5;
                              S8;
                         ELSE S6;
                   S9;
              ELSE S1;
        S10;
```



بدیهی است که محاسبه پیچیدگی های گراف های کنترل فوق با عدد سیکلوماتیک کلاسیک می دهد:

$$v(a) = v(b) = 3; \; v(c) = v(d) = v(e) = 4.$$

اکنون امکان تمایز بین پیچیدگی برنامه های الف و ب وجود ندارد. از این روی ، اگر بخواهیم گراف های کنترل دارای پیچیدگی های یکسان که با عدد سیکلوکلیماتیک مک کیب اندازه گیری می شوند را متمایز کنیم، بایستی یک مفهوم جدید را در نظر بگیریم که در بخش زیر معرفی خواهد شد.

2- درجه تودرتویی

فرض کنید کهP یک برنامه یا یک واحد(ماژولی) از یک برنامه باشد. گراف جهت دار P با $G = (V,A),$ نعریف می شود که v، یک مجموعه غیر تهی متناهی می باشد و A یک رابطه در V است.هر درایه در V موسوم به یک گره بوده و هر زوج در A موسوم به قوس یا یال است. سپس

$$V = V_1 + V_2$$

که

$$V = \{x \mid x \in V, \mid suc(x) \mid \leq 1\};$$

$$V = \{x \mid x \in V, \mid suc(x) \mid \geq 2\};$$

$suc(x)$ بیانگر مجموعه ای از توالی های گره X(X گره انتهایی نیست)،، $suc(x) = \{y \mid (x,y) \in A, x,y \in V\};$ ، $\mid suc(x) \mid$ بیانگر درجه خروجی گره های X(تعداد درایه ها یا مولفه های $suc(x)$) است.

مولفه های V_2 موسوم به گره های تصمیم یا گره های انتخاب هستند.(یادداشت: قوس بین X وy، گره X را ترک کرده و وارد گره Y می شود). نظر ما این است که x بر y مقدم بوده و y تالی x است).

هر گونه ارجاع به گراف در ادامه این مقاله مربوط به گراف همبند جهت دار با یک گراف اولیه است(یک گره که فاقد گره اسبق است).

در زیر، ما به توصیف رابطه غالبیت پیشرو می پردازیم که در یک گراف جهت دار و البته در تحلیل جریان کنترل وجود دارد(برای تعریف غالبیت پسین یا پیش غالبیت، به HECH77، صفحه 55). فرض کنید که s گره اولیه و e گره خروجی کراف G باشد.

تعریف: اگر $x, y \in V, x \neq y$ باشد، آنگاه X بر y غالب پیشرو یا پیش غالب خواهد بود اگر و تنها اگر هر مسیر در G از Y به گروه خروجی آن حاوی x است. یادداشت:x بر y پیش غالب است. به طور ساده تر، فرض کنید که $PDOM(y) = \{x \mid x \text{ pdom } y, x \in V\}$ به ازای $y \in V$ باشد. گفته می شود که اگر و تنها اگر 1- x بر y غالب باشد و 2 – اگر $z \in V$، z بر y غالب باشد، آنگاه z بر x غالب است، x مستقیما بر y غالب خواهد بود.

مثال: گراف برنامه 1 را با $s = 1$ و $e = 7$ در نظر بگیرید. سپس، $V = \{1, 2, 3, 4, 5, 6, 7\}, V_1 = \{2, 3, 5, 6, 7\}, V_2 = \{1, 4\}$، 1 بر 1، 4 بر 1، 7 بر 1 و 4 مستقیما بر 1 غالب است.

اثبات قضیه زیر در پیوست ارایه شده است(به نتیجه 5 مراجعه کنید).

قضیه:اگر $x \in V_2$ باشد، آنگاه x دارای یک گره کنترلی پیشرو مستقیم x' است

اکنون این گره کنترلی پیشرو x' از یک مولفه x از V در نظر گرفته می شود(این مورد در قضیه فوق وجود داشته و منحصر به فرد است). گره ها و یال های بین x و x'، زیرگراف $G' \subset G$ را می سازد، G'موسوم به زیر گراف کنترل گره x است. بدیهی است که G' دارای تنها یک گره ورودی (x) و یک گروه خروجی (x') می باشد. فرض کنید که $N = |V_2| (|V_2|$ بیانگر تعداد مولفه های v1 است، $n \in V_2$, و G_n، زیر گراف کنترل گره n است.

فرض کنید که d_n تعداد گره های انتخاب G است.

سپس، درجه تو درتویی گره انتخاب n به صورت زیر تعریف می شود

$$\epsilon_n = 1 - (1/d_n)$$

از آن جا که G_n دارای یک گره انتخاب n است از این روی $d_n \geq 1$, و $0 \leq \epsilon_n < 1$است.

اکنون درجه تودرنویی یک گراف G را به صورت زیر تعریف می کنیم

$$\epsilon = (\epsilon_1 + \epsilon_2 + \ldots + \epsilon_N)/N$$

بدیهی است که $0 \leq \epsilon < 1$

3- اصلاح **v(G)**

ما یک شاخص پیچیدگی جدید را برای برنامه ها به صورت زیر تعریف می کنیم

$$c(G) = v(G) + \epsilon$$

که $\mathbf{v(G)}$ یک عدد سیکلوماتیک بوده و درجه تودرتویی G است که در بالا تعریف شده است.

شاخص پیچیدگی مک کیب تنها ساختار تصمیم یک برنامه را در نظر گرفته و مستقل از ساختار تو درتو است.

اگرچه برای مثال، $\mathbf{v(G)}$ برای دو برنامه متفاوت مشابه است: احتمال دارد که آن ها دارای یک درجه متفاوتی از تودرتویی باشند با این حال، عدد سیکلوماتیک آن ها را متمایز نمی کند. با این حال، افزایش درجه تودرتویی مستلزم اصلاح شاخص پیچیدگی مک کیب است.

محاسبه پیچیدگی های جدید برنامه های $\mathbf{(a)}$-$\mathbf{(c)}$ به صورت زیر، میدهد:

$$
\begin{array}{ll}
\text{1)} & d_1 = d_2 = 1, \; so \;\; \epsilon_1 = \epsilon_2 = 0, \epsilon = 0, c(a) = v(a) = 3; \\
\text{2)} & d_1 = 2, d_2 = 1, \; so \;\; \epsilon = 1/4, c(b) = 3{,}25; \\
\text{3)} & d_1 = d_2 = d_3 = 1, \; so \;\; \epsilon = 0, c(c) = v(c) = 4; \\
\text{4)} & d_1 = 3, d_2 = d_3 = 1, \; so \;\; \epsilon = 1/9, c(d) = 4{,}11; \\
\text{5)} & d_1 = 3, d_2 = 2, d_3 = 1, \; so \;\; \epsilon = 5/18, c(e) = 4{,}28.
\end{array}
$$

اکنون امکان تمایز بین برنامه های a-b،(با یک عدد سیکلوماتیک) و بین برنامه های c-d-e وجود دارد (که دارای عدد سیکلوماتیک یکسان است)

$$\mathrm{c(a)} < \mathrm{c(b)} \text{ and } \mathrm{c(c)} < \mathrm{c(d)} < \mathrm{c(e)}.$$

مثال : برای دوازده نمونه از مک کیب(MCCA76، صفحه 310)، عدد سیکلوماتیک $\mathbf{v(G)}$و شاخص پیچیدگی تعمیم یافته $\mathbf{c(G)}$ محاسبه شده و بر اساس رتبه ان ها مقایسه می شود

Graph G (no.)	v(G)	c(G)	Rank of v(G)	Rank of c(G)
1	2	2	1	1
2	3	3,25	2	2
3	5	5,50	3	3
4	6	6,55	4	4
5	8	8,39	5,5	5
6	8	8,60	5,5	6
7	9	9,46	7	7
8	10	10,48	9	10
9	10	10,17	9	9
10	10	10,09	9	8
11	11	11,56	11	11
12	19	19,42	12	12

5- پیوست

فرض کنید که $G = (V,A,s,e)$ یک گراف کنترل برنامه است که در آن v بیانگر مجموعه ای از همه گره هاست،A مجموعه ای از همه یال ها بوده، S گره اولیه و e گره خروجی است. هیچ گونه کاهش عمومیت با فرض این که G دارای دقیقا یک گره اولیه و یک گروه خروجی است وجود ندارد

تعریف: در صورتی که $x, y \in V, x \neq y$، باشد، آنگاه x بر y غالب است اگر و تنها اگر هر مسیر در G ازY به گروه خروجی آن، حاوی X است. یادداشت: X بر Y غالب است. به طور ساده تر، فرض کنید که $PDOM(y) = \{x \mid x \text{ pdom } y, x \in V\}$ به ازای هر $y \in V$ غالب است.

می توان گفت که x بر y مستقیم غالب است اگر و تنها اگر 1- x بر y غالب باشد و 2- اگر $z \in V$، z بر y غالب باشد، آنگاه z بر x غالب خواهد بود

پنج نتیجه گیری برای رابطه پسا غالبیت postdominance وجود دارد

نتیجه 1: $PDOM(e) = \{e\}$

اثبات : بدیهی

نتیجه 2: رابطه پساغالبیت یک گراف G، ترتیب جزیی است.

یادداشت: یک ترتیب جزیی در مجموعه S، یک رابطه انعکاسی ، غیر متقارن و انتقالی بر S است)

اثبات:رابطه انعکاسی: به ازای هر $x \in V$، از تعریف می توان دریافت که X برX غالب است

2- رابطه غیر متقارن : ادعای ما این است که اگر $x, y \in V$، x بر y غالب باشد و y بر x غالب باشد، آنگاه $x = y$ است. اولا، اگر $x = e$ یا $y = e$، بر اساس نتیجه 1 می توان گفت که $x = y = e$، است.اکنون دومین مورد را در نظر می گیریم $x \neq e$ و $k < y$ به طوری که $wi_1 = x$، است زیرا y بر xغالب است، و از این روی $j_1, i_1 < j_1 < k,$ وجود داردبه طوری که $wj_1 = y$ است. ما این کار را ادامه می دهیم. در نهایت $i_1, i_2, \dots, j_1, j_2, \dots, 1 < i_1 < j_1 < i_2 < j_2 < \dots < k,$ وجود دارد به طوری که $wi_1 = wi_2 = \dots = x, wj_1 = wj_2 = \dots = y$ است. زیرا k یک عدد متناهی است.

3- رابطه تراگذاری: اگر x بر y و y بر z غالب باشد آنگاه x برz غالب خواهد بود(بدیهی)

نتیجه 3: گره خروجی e از G بر همه گره ها غالب است.

اثبات: بدیهی

نتیجه 4: اگر $x \quad V$، آنگاه $PDOM(x)$ ایجاد یک زنجیره می کند(یادداشت: یک زنجیره یا ترتیب خطی بر روی مجموعه S ترتیب جزیی درS می باشد به طوری که هر جفت از مولفه ها قابل مقایسه هستند.

اثبات: به ازای $x \quad V$، $PDOM(x)$ یک ترتیب جزیی است. اکنون می توان نشان داد که هر جفت از مولفه های $PDOM(x)$ با رابطه غالبیت قابل مقایسه است. این برای $y, z \in PDOM(x)$ به این معنی است که یا Y بر Z و یا Z بر Y غالب است.فرض کنید که $(w1, w2, \dots, wk)$ یک مسیر غیر گردشی با $w1 = x, wk = e$، است. سپس، برخی شاخص ها $i, j, 1 \leq i, j \leq k,$ وجود دارد به طوری که $wi = y, wj = z$ باشد. ما ادعا می کنیم که Z برY غالب است(Y بر X غالب است). فرض کنید کهZ بر Yغالب نباشد. آنگاه یک مسیر P از Y به

e وجود دارد که فاقد z است. اکنون $(w1, \ldots, wi) + p$ یک مسیر از x به e است که فاقد z می باشد. با این حال این ناقض فرض $z \in PDOM(x)$، است.

نتیجه 5 : هر گره به جز e دارای یک گره کنترلی مستقیم و منحصر به فرد است

اثبات: فرض کنید که x یک گره و $x \neq e$ باشد. آنگاه $\{x,e\} \subset PDOM(x)$ است. بر اساس نتیجه 4، رابطه پسا غالبیتف یک زنجیره در $PDOM(x)$ و نیز بر روی $PDOM(x) - \{x\}$ است. می دانیم که هر مجموعه غیر تهی متناهی دارای یک مولفه حداقل و حداکثر است. $PDOM(x) - \{x\}$ یک زنجیره غیر تهی متناهی است. از این روی، دارای حداقل یک مولفه منحصر به فرد است که بایستی مستقیما بر **x** غالب باشد.