دانشکده آموزش‌های الکترونیکی

پایان‌نامه کارشناسی ارشد در رشته مهندسی فناوری اطلاعات (تجارت الکترونیک)

پیش بینی خطای نرم‌افزار بر اساس وابستگی‌های قطعات کد با استفاده از داده‌کاوی

به کوشش:

فرزانه مالکی

استاد راهنما:

جناب آقای دکتر اشکان سامی

دی ۱۳۹۰

بسم الله الرحمن الرحيم

به نام خدا

اظهار نامه

اینجانب فرزانه مالکی دانشجوی رشته مهندسی فناوری اطلاعات گرایش تجارت الکترونیک دانشکده آموزش‌های الکترونیکی اظهار می‌نمایم که این پایان‌نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌کنم که تحقیق و موضوع پایان‌نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین‌نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام‌خانوادگی

تاریخ و امضا

به نام خدا

پیش بینی خطای نرم‌افزار بر اساس وابستگی‌های قطعات کد با استفاده از داده‌کاوی

به کوشش

فرزانه مالکی

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه شیراز به عنوان بخشی

از فعالیت‌های تحصیلی لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی :

مهندسی فناوری اطلاعات (گرایش تجارت الکترونیک)

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی کمیته پایان نامه، بادرجه‌ی: .....

دکتر اشکان سامی، استادیار دانشکده برق و کامپیوتر (رئیس کمیته) .....

دکتر امید بوشهریان، استادیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز .....

دکتر شهریار لطفی، استادیار گروه علوم کامپیوتر، دانشگاه تبریز .....

دی ماه ۱۳۹۰

تقدیم به اولین معلمان زندگی

پدر

مادر

آموزگارانی که برایم زندگی؛ بودن و انسان بودن را معنا کردند

و به همسرم

که بودنش، باورش و حمایتش

به من توان دوباره روییدن داد

سپاسگزاری

حال که با لطف و عنایت خداوند متعال رساله حاضر به پایان رسیده است بر خود لازم می‌دانم که از استاد ارجمند آقای دکتر اشکان سامی به خاطر زحمات فراوان ایشان در راه به ثمر رساندن این تحقیق تشکر و قدردانی نمایم.

چکیده

پیش بینی خطای نرم‌افزار بر اساس وابستگی‌های قطعات کد با استفاده از داده‌کاوی

به کوشش

فرزانه مالکی

تضمین کیفیت نرم‌افزار یکی از چالش برانگیزترین مباحث تولید نرم‌افزار می‌باشد و هدف آن تولید نرم‌افزاری بدون خطا است که نیاز و انتظارات مشتری را پاسخگو باشد. یکی از فاکتورهایی که منجر به تعداد بالای خطاهای نرم‌افزاری می‌شود اینست که منابع تضمین کیفیت توسط زمان و هزینه محدود می‌شوند. به منظور اختصاص منابع به صورت موثر، مدیران بایستی منابع را به بخش‌هایی از محصول نرم‌افزاری اختصاص دهند که احتمال معیوب بودن بیشتری نسبت به سایر بخش‌ها دارند. مدیران برای شناسایی قسمت‌های معیوب کد می‌توانند از روش‌های خودکار پیش‌بینی خطا استفاده کنند که معمولا بر اساس معیارهای کد و وابستگی‌هایی که بین اجزای کد وجود دارند عمل می‌کنند. تلاش این رساله ارائه معیاری کارا بر اساس وابستگی‌های موجود بین اجزای کد می‌باشد. در این رساله ما به توسعه مفهوم وابستگی چرخشی در گراف وابستگی کد پرداخته و با ارزیابی معیار مورد نظر بر روی چندین پروژه متن باز و مقایسه این معیار با معیارهای کد نشان داده‌ایم که وابستگی چرخشی عملکرد بهتری نسبت به معیارهای کد در پیش‌بینی خطا دارد

فهرست

عنوان **صفحه**

۱-مقدمه..... ۱

۲-پیش‌زمینه و بررسی مطالعات گذشته..... ۳

۲-۱- معیارهای محصول..... ۳

۲-۲- معیارهای وابستگی..... ٤

۲-۳- معیارهای تاریخی..... ٥

۲-٤- بررسی فعالیت‌های گذشته..... ٥

۲-۴-۱- بررسی فعالیت‌ها در زمینه معیارهای محصول ٥

۲-۴-۲- بررسی فعالیت‌ها در زمینه معیارهای تاریخی..... ۸

۲-۴-۳- بررسی فعالیت‌ها در زمینه معیارهای وابستگی..... ۱۰

۳- وابستگی چرخشی..... ۱۲

۴- جمع آوری داده‌ها:..... ۱۸

۵- نتایج..... ۲۸

۵-۱- بررسی نتایج نرم‌افزار تامکت..... ۷۳

۵-۲- بررسی اکلیپس ۲,۰..... ۷۵

۵-۳- بررسی اکلیپس ۲,۱..... ۷۵

۵-۴- بررسی اکلیپس ۳,۰..... ۷٦

۶- نتیجه‌گیری..... ۷۸

مراجع..... ۷۹

فهرست شکل‌ها

عنوان **صفحه**

شکل شماره ۱- گراف وابستگی ۱۵
شکل شماره ۲- وابستگی چرخشی نوع اول ۱۶
شکل شماره ۳- وابستگی چرخشی نوع دوم ۱۶
شکل شماره ۴- وابستگی چرخشی نوع سوم ۱۶
شکل شماره ۵- وابستگی چرخشی نوع چهارم ۱۷
شکل شماره ۶- محیط کلی ابزار CDA ۲۲
شکل شماره ۷ -نمایش وابستگی‌ها به صورت گراف در ابزار CDA ۲۳
شکل شماره ۸- نمایش وابستگی‌ها در ابزار CDA ۲۳
شکل شماره ۹-گراف ۲۴
شکل شماره ۱۰- ماتریس وابستگی ۲۵
شکل شماره ۱۱- مقادیر پیش‌بینی شده و مشاهده شده ۳۰
شکل شماره ۱۲- منحنی ROC ۳۳

فهرست جدول‌ها

عنوان **صفحه**

جدول شماره ۱- مشخصات داده‌های رساله ۲۱
جدول شماره ۲ -معیارهای استخراج شده از برنامه پیرست ۲۶
جدول شماره ۳- معیارهای زیمرمن ۲۷
جدول شماره ۴- دسته‌کننده‌های مورد استفاده قرار گرفته شده در رساله ۳۴
جدول شماره ۵- صحت و انحراف معیار در نرم افزار تامکت ۳۸
جدول شماره ۶- صحت و انحراف معیار در نرم افزار اکلیپس ۲,۰ ..... ۴۱
جدول شماره ۷- صحت و انحراف معیار در نرم افزار اکلیپس ۲,۱ ..... ۴۵
جدول شماره ۸- صحت و انحراف معیار در نرم افزار اکلیپس ۳,۰ ..... ۴۸

فهرست نمودارها

عنوان **صفحه**

نمودار شماره ۱- مقایسه اندازه فراخوانی در نرم‌افزار تامکت ۳۵
نمودار شماره ۲- مقایسه اندازه دقت در نرم‌افزار تامکت ۳۵
نمودار شماره ۳- مقایسه اندازه معیار اف در نرم‌افزار تامکت ۳۶
نمودار شماره ۴- مقایسه اندازه صحت در نرم‌افزار تامکت ۳۶
نمودار شماره ۵- مقایسه اندازه ناحیه زیر منحنی ROC در نرم‌افزار تامکت ۳۷
نمودار شماره ۶- مقایسه اندازه انحراف معیار در نرم‌افزار تامکت..... ۳۷
نمودار شماره ۷ - مقایسه اندازه فراخوانی در نرم‌افزار اکلیپس ۲,۰ ..... ۳۸
نمودار شماره ۸- مقایسه اندازه دقت در نرم‌افزار اکلیپس ۲,۰ ..... ۳۹
نمودار شماره ۹- مقایسه اندازه معیار اف در نرم‌افزار اکلیپس ۲,۰ ..... ۳۹
نمودار شماره ۱۰- مقایسه اندازه صحت در نرم‌افزار اکلیپس ۲,۰ ..... ٤۰
نمودار شماره ۱۱- مقایسه اندازه ناحیه زیر منحنی ROC در نرم‌افزار اکلیپس ۲,۰ ..... ٤۰
نمودار شماره ۱۲- مقایسه اندازه انحراف معیار در نرم‌افزار اکلیپس ۲,۰..... ٤۱
نمودار شماره ۱۳- مقایسه اندازه فراخوانی در نرم‌افزار اکلیپس۲,۱ ..... ٤۲
نمودار شماره ۱۴- مقایسه اندازه دقت در نرم‌افزار اکلیپس۲,۱ ..... ٤۲
نمودار شماره ۱۵- مقایسه اندازه معیار اف در نرم‌افزار اکلیپس۲,۱ ..... ٤۳
نمودار شماره ۱۶- مقایسه اندازه صحت در نرم‌افزار اکلیپس۲,۱..... ٤۳
نمودار شماره ۱۷- مقایسه اندازه ناحیه زیر منحنی ROC در نرم‌افزار اکلیپس۲,۱ ..... ٤٤
نمودار شماره ۱۸- مقایسه اندازه انحراف معیار در نرم‌افزار اکلیپس۲,۱..... ٤٤
نمودار شماره ۱۹- مقایسه اندازه فراخوانی در نرم‌افزار اکلیپس۳,۰ ..... ٤٥
نمودار شماره ۲۰- مقایسه اندازه دقت در نرم‌افزار اکلیپس۳,۰ ..... ٤٦
نمودار شماره ۲۱- مقایسه اندازه معیار اف در نرم‌افزار اکلیپس۳,۰ ..... ٤٦

عنوان **صفحه**

نمودار شماره ۲۲- مقایسه اندازه صحت در نرم‌افزار اکلیپس۳,۰ ..... ٤٧
نمودار شماره ۲۳- مقایسه اندازه ناحیه زیر منحنی ROC در نرم‌افزار اکلیپس۳,۰ ..... ٤٧
نمودار شماره ۲۴- مقایسه اندازه انحراف معیار در نرم‌افزار اکلیپس۳,۰..... ٤٨
نمودار شماره ۲۵- مقایسه معیار فراخوانی با استفاده از دسته‌کننده Bayesian Logistic Regression ٤٩
نمودار شماره ۲۶- مقایسه معیار دقت با استفاده از دسته‌کننده Bayesian Logistic Regression ٤٩
نمودار شماره ۲۷- مقایسه معیار صحت با استفاده از دسته‌کننده Bayesian Logistic Regression ٥٠
نمودار شماره ۲۸- مقایسه معیار اف با استفاده از دسته‌کننده Bayesian Logistic Regression ٥٠
نمودار شماره ۲۹- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده Bayesian Logistic Regression ٥١
نمودار شماره ۳۰- مقایسه نحراف معیار با استفاده از دسته‌کننده Bayesian Logistic Regression..... ٥١
نمودار شماره ۳۱- مقایسه معیار فراخونی با استفاده از دسته‌کننده Bayes Net ٥٢
نمودار شماره ۳۲ -مقایسه معیار دقت با استفاده از دسته‌کننده Bayes Net ٥٢
نمودار شماره ۳۳ -مقایسه معیار اف با استفاده از دسته‌کننده Bayes Net ٥٣
نمودار شماره ۳۴- مقایسه معیار صحت با استفاده از دسته‌کننده Bayes Net..... ٥٣
نمودار شماره ۳۵- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کنندهBayes Net..... ٥٤
نمودار شماره ۳۶ - مقایسه انحراف معیار با استفاده از دسته‌کنندهBayes Net ٥٤
نمودار شماره ۳۷- مقایسه معیار فراخوانی با استفاده از دسته‌کنندهNaive Bayes..... ٥٥
نمودار شماره ۳۸ -مقایسه معیار دقت با استفاده از دسته‌کنندهNaive Bayes..... ٥٥
نمودار شماره ۳۹- مقایسه معیار اف با استفاده از دسته‌کنندهNaive Bayes..... ٥٦
نمودار شماره ۴۰ -مقایسه معیار صحت با استفاده از دسته‌کنندهNaive Bayes..... ٥٦
نمودار شماره ۴۱- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کنندهNaive Bayes..... ٥٧
نمودار شماره ۴۲- مقایسه انحراف معیار ROC با استفاده از دسته‌کنندهNaive Bayes..... ٥٧
نمودار شماره ۴۳ -مقایسه معیار فراخوانی با استفاده از دسته‌کننده Logistic ٥٨
نمودار شماره ۴۴ -مقایسه معیار دقت با استفاده از دسته‌کننده Logistic ٥٨
نمودار شماره ۴۵- مقایسه معیار اف با استفاده از دسته‌کننده Logistic ٥٩
نمودار شماره ۴۶ -مقایسه معیار صحت با استفاده از دسته‌کننده Logistic ٥٩
نمودار شماره ۴۷- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده Logistic ٦٠
نمودار شماره ۴۸- مقایسه انحراف معیار با استفاده از دسته‌کننده Logistic ٦٠
نمودار شماره ۴۹- مقایسه معیار فراخوانی با استفاده از دسته‌کننده Simple Logistic ٦١
نمودار شماره ۵۰ -مقایسه معیار دقت با استفاده از دسته‌کننده Simple Logistic ٦١
نمودار شماره ۵۱- مقایسه معیار اف با استفاده از دسته‌کننده Simple Logistic ٦٢
نمودار شماره ۵۲- مقایسه معیار صحت با استفاده از دسته‌کننده Simple Logistic ٦٢

عنوان **صفحه**

نمودار شماره ۵۳- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده Simple Logistic ٦٣
نمودار شماره ۵۴- مقایسه انحراف معیار با استفاده از دسته‌کننده Simple Logistic ٦٣
نمودار شماره ۵۵ - مقایسه معیار فراخوانی با استفاده از دسته‌کننده Classification via regression..... ٦٤
نمودار شماره ۵۶- مقایسه معیار دقت با استفاده از دسته‌کننده Classification via regression..... ٦٤
نمودار شماره ۵۷- مقایسه معیار اف با استفاده از دسته‌کننده Classification via regression..... ٦٥
نمودار شماره ۵۸- مقایسه معیار صحت با استفاده از دسته‌کننده Classification via regression..... ٦٥
نمودار شماره ۵۹- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده Classification via regression..... ٦٦
نمودار شماره ۶۰- مقایسه انحراف معیار با استفاده از دسته‌کننده Classification via regression..... ٦٦
نمودار شماره ۶۱- مقایسه معیار فراخوانی با استفاده از دسته‌کننده Decision Table..... ٦٧
نمودار شماره ۶۲- مقایسه معیار دقت با استفاده از دسته‌کننده Decision Table..... ٦٧
نمودار شماره ۶۳- مقایسه معیار اف با استفاده از دسته‌کننده Decision Table..... ٦٨
نمودار شماره ۶۴- مقایسه معیار صحت با استفاده از دسته‌کننده Decision Table..... ٦٨
نمودار شماره ۶۵- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده Decision Table..... ٦٩
نمودار شماره ۶۶- مقایسه انحراف معیار با استفاده از دسته‌کننده Decision Table..... ٦٩
نمودار شماره ۶۷- مقایسه معیار فراخوانی با استفاده از دسته‌کننده J۴۸..... ٧٠
نمودار شماره ۶۸- مقایسه معیار دقت با استفاده از دسته‌کننده J۴۸..... ٧٠
نمودار شماره ۶۹- مقایسه معیار اف با استفاده از دسته‌کننده J۴۸..... ٧١
نمودار شماره ۷۰- مقایسه معیار صحت با استفاده از دسته‌کننده J۴۸..... ٧١
نمودار شماره ۷۱- مقایسه ناحیه زیر منحنی ROC با استفاده از دسته‌کننده J۴۸..... ٧٢
نمودار شماره ۷۲ -مقایسه انحراف معیار با استفاده از دسته‌کننده J۴۸..... ٧٢

۱- مقدمه

تضمین کیفیت[۱] نرم‌افزار یکی از چالش برانگیزترین مباحث تولید نرم‌افزار[۲] می‌باشد. سازمان ملی استاندارد در سال ۲۰۰۲ اعلام کرد خطاهای نرم‌افزاری سالیانه خسارتی در حدود ۶۰ میلیارد دلار به اقتصاد آمریکا وارد می‌نماید.(Tassey, ۲۰۰۲) تضمین کیفیت از ابتدای پروژه‌ی تولید نرم‌افزار آغاز می‌شود و هدف آن تولید نرم‌افزاری بدون خطا، طبق زمانبندی و بودجه مشخص شده می‌باشد که نیاز و انتظارات مشتری را پاسخگو باشد. یکی از فاکتورهایی که منجر به تعداد بالای خطاهای نرم‌افزاری در نتیجه تولید نرم‌افزاری با کیفیت پایین می‌شود محدودیت منابع تضمین کیفیت می‌باشد. این منابع توسط زمان و هزینه محدود می‌شوند از اینرو مدیران تیم‌های تضمین کیفیت در تلاشند که منابع را به بخش‌هایی از محصول نرم‌افزاری اختصاص دهند که احتمال معیوب بودن بیشتری نسبت به سایر بخش‌ها دارند. اما سوال اینجاست که اجزایی که احتمال معیوب بودن آنها بالاست کدامند؟ در اینجا مبحثی به نام پیش‌بینی خطا[۳] مطرح می‌شود که هدف آن طرح ایده‌هایی برای الویت‌بندی در شناسایی خطا می‌باشد. مدیران تیم‌های تست نرم‌افزار با استفاده از یک مدل پیش‌بینی خطای[۴] مناسب می‌توانند منابع را به طور موثری به منظور شناسایی خطا اختصاص دهند. ساخت مدل‌های پیش‌بینی خطا به روش‌های مختلف انجام می‌شود. یکی از این روش‌ها که امروزه کاربرد فراوانی دارند روش‌های خودکار می‌باشند. داده‌کاوی و تکنیک‌های یادگیری ماشین یکی از موفق‌ترین روش‌های خودکار پیش‌بینی خطا می‌باشد.

---

۱ quality assurance
۲ software
۳ defect prediction
۴ predictor model

به طور معمول روش‌های خودکار پیش‌بینی خطا با استفاده از معیارهای نرم‌افزار صورت می‌گیرد. معیارهای نرم‌افزاری مشخصات و ویژگی‌های کد را اندازه‌گیری می‌نمایند. مطالعات زیادی ارتباط معیارهای نرم‌افزاری با خطاهای نرم‌افزاری را نشان می‌دهد. تحقیقات نشان می‌دهد علاوه بر معیارهای نرم‌افزاری، معیارهایی که بر اساس اطلاعات تاریخی یک محصول و همچنین وابستگی‌های موجود بین اجزای کد محاسبه می‌شوند نیز در پیش‌بینی خطاهای نرم‌افزاری موفق هستند.

همان‌طور که گفته شد یکی روش‌هایی که در پیش‌بینی خطا مطرح شده است، مساله مربوط به وابستگی[۵]های بین اجزای کد می‌باشد(Zimmermann & Napappan, ۲۰۰۸) ، (Binkley & Schach, ۱۹۹۸) (Nagappan, Ball, & Zeller, ۲۰۰۶) (Zimmermann, Premraj, & Zeller, ۲۰۰۷) انواع وابستگی‌ها می‌توانند بین دو جز کد (کلاس[۶]، ماژول[۷]، فایل[۸]...) مطرح شوند که تحقیقات نشان داده است این وابستگی‌ها ارتباط بسیار زیادی با وجود خطاهای مفهومی[۹] دارند. در این رساله معیاری بر اساس وابستگی موجود بین اجزای کد تعریف می‌شود که با عنوان وابستگی چرخشی[۱۰] از آن نام می‌بریم. وابستگی چرخشی از تعمیم وابستگی خروجی[۱۱] که از معیارهای موفق در پیش‌بینی خطا می‌باشد به دست می‌آید.(Kafura & Henry, ۱۹۸۱) در ادامه با استفاده از تکنیک‌های داده‌کاوی به بررسی چگونگی عملکرد وابستگی چرخشی در مقایسه با سایر معیارها می‌پردازیم.

ساختار این رساله بدین شرح است: در فصل دوم به بررسی پیش‌زمینه و تلاش‌های گذشته می‌پردازیم. فصل سوم وابستگی چرخشی معرفی می‌شود. جمع آوری داده در فصل چهارم مطرح می‌گردد. فصل پنجم شامل نتایج و آزمایشات رساله می‌باشد و فصل آخر، فصل ششم بیانگر نتیجه‌گیری است.

---

۵ dependency
۶ class
۷ module
۸ file
۹ semantic errors
۱۰ circular dependency
۱۱ fan out

۲- پیش‌زمینه و بررسی مطالعات گذشته

تلاش‌های گذشته با هدف پیش‌بینی خطا را در سه گروه مورد بررسی قرار می‌دهیم: معیارهای محصول[۱۲]، معیارهای وابستگی[۱۳] و معیارهای تاریخی[۱۴]. ابتدا به بیان توضیحی در مورد شیوه کار کردن این معیارها می‌پردازیم

۲-۱- معیارهای محصول

معیارهای نرم‌افزار[۱۵] برای اندازه‌گیری درجه پیشرفته بودن یک محصول و یا یک فرآیند نرم‌افزاری به کار می‌روند. معیارهای نرم‌افزاری به چندین گروه تقسیم می‌شوند: معیارهای محصول، معیارهای فرآیند[۱۶]، معیارهای پروژه[۱۷] و معیارهای منبع[۱۸]. معیارهای نرم‌افزاری که در پیش‌بینی خطا به کار می‌روند معیارهای محصول می‌باشند که از مشخصات کد سیستم نرم‌افزاری استخراج می‌شوند. این معیارها به سه گروه تقسیم می‌شوند: معیارهای اندازه[۱۹]، معیارهای پیچیدگی [۲۰] و

---

۱۲ product metrics
۱۳ dependency metrics
۱۴ historical metrics
۱۵ software metrics
۱۶ process metrics
۱۷ project metrics
۱۸ resource metrics
۱۹ size metrics
۲۰ complexity metrics

معیارهای کیفیت[۲۱].(Mills, ۱۹۸۸) معیارهای اندازه بر اساس تعداد خطوط کد برنامه محاسبه می‌شوند مانند تعداد کل خطوط برنامه[۲۲]، تعداد خطوط توضیحات[۲۳]و.... معیارهای پیچیدگی مستقیما به میزان نگهداشت‌پذیری[۲۴] و قابلیت تست برنامه وابسته است از جمله معروف‌ترین معیارهای پیچیدگی معیارهای پیچیدگی مک کیب[۲۵] و معیارهای هالستد[۲۶] می‌باشند. معیارهای مک‌کیب پیچیدگی کد را بر اساس تعداد مسیرهای کنترلی محاسبه می‌نماید.(McCabe, ۱۹۷٦) هالستد معیارهای خود را بر اساس ارتباطات ریاضی بین اجزای کد، پیچیدگی کد و نوع زبان برنامه‌نویسی مطرح کرد. (Halstead, ۱۹۷۵) معیارهای اتصال[۲۷] و پیوستگی[۲۸] از معروف‌ترین معیارهای کیفیت می‌باشند که بالا و یا پایین بودن اندازه این دو معیار نشان دهنده کیفیت محصول و یا فرآیند نرم‌افزاری است.(Pressman, ۱۹۸۲) معیارهای کیفیت معیارهایی می‌باشند که درجه آنها می‌تواند تولیدکنندگان نرم‌افزار را در مورد توانایی دست کار کردن سیستم‌شان مطمئن سازد.

۲-۲- معیارهای وابستگی

معیارهای وابستگی معیارهایی هستند که بر اساس ارتباط اجزای کد نرم‌افزار محاسبه می‌شوند. این ارتباط می‌تواند بین سطوح مختلف کد مطرح شود مانند سطح فایل، کلاس، تابع..... نوع ارتباط و یا به عبارت دیگر وابستگی موجود بین اجزای کد نیز می‌تواند متفاوت باشد مانند وابستگی داده[۲۹]

[۲۱] quality metrics
[۲۲] total line of code
[۲۳] line of code comment
[۲٤] maintainability
[۲٥] McCabe metrics
[۲٦] Halstead metrics
[۲۷] coupling
[۲۸] cohesion
[۲۹] data dependency

که بر اساس تعریف و استفاده از داده می‌باشد و یا وابستگی صدا زدن[۳۰] که بر اساس تعریف و صدا زدن مولفه‌ها می باشد.(Zimmermann & Nagappan, ۲۰۰۸)

۲-۳- معیارهای تاریخی

این دسته از معیارهای بر اساس تغییراتی که در بین چندین انتشار[۳۱] مختلف از یک سیستم نرم‌افزاری رخ می‌دهد تعریف می‌شوند. این تغییرات می توانند اضافه شدن، حذف شدن، تغییر یافتن مولفه‌های کد در بین چندین انتشار باشند. محدودیتی که در به کار بردن این معیارها وجود دارد این است که محاسبه این معیارهای تنها برای محصولات نرم‌افزاری مقدور می‌باشد که دارای چندین انتشار باشند و همچنین اطلاعات کامل در مورد انتشارات قبلی موجود باشد.

۲-۴- بررسی فعالیت‌های گذشته

در ادامه به بررسی تلاش‌های گذشته در زمینه پیش‌بینی خطا می‌پردازیم.

۲-۴-۱- بررسی فعالیت‌ها در زمینه معیارهای محصول

مطالعاتی که بر اساس معیارهای کد صورت گرفت بدین شرح می‌باشند:

---

[۳۰] call dependency
[۳۱] release

در اولین مطالعات در زمینه پیش‌بینی خطا، آکیاما در سال ۱۹۷۱ بر روی یک سیستم که در فوجیتسو ژاپن توسعه یافته بود کار کرد و نشان داد معیارهای پیچیدگی نرم‌افزار و معیارهای اندازه نرم‌افزار در پیش‌بینی خطا موفق عمل می‌نمایند.(Akiyama, ۱۹۷۱) فردینند در سال ۱۹۷۴ ارتباط تعداد خطا با تعداد اجزای کد را بررسی نمود و به این نتیجه رسید که تعداد اجزای کد با تعداد خطا در ارتباط است.(Ferdinand, ۱۹۷٤) در سال ۱۹۷۵هالستد معیارهای نرم‌افزاری را ارائه داد که بر اساس پیچیدگی زبان برنامه‌نویسی عمل می‌نمودند. هالستد از این معیارها برای پیش‌بینی خطا استفاده کرد و نشان داد معیارهای پیچیدگی هالستد در پیش‌بینی خطا موفق عمل می‌نمایند. این معیارها با نام معیارهای هالستد شناخته می‌شوند. (Halstead, ۱۹۷٥) چیدمبر و کمرر در سال ۱۹۹۴ بر روی سیستم‌های شی‌گرا کار کردند و شش معیار طراحی برای این نوع سیستم معرفی کردند. این معیارها با نام معیارهای طراحی چیدمبر و کمرر شناخته می‌شوند. (Chidamber & Kemerer, ۱۹۹٤) باسیلی در سال ۱۹۹۶ از معیارهای چیدمبر و کمرر به منظور پیش‌بینی خطا استفاده نمود و نشان داد معیارهای چیدمبر و کمرر عملکرد موفقی در پیش‌بینی خطا دارند.(Basili, Briand, & Melo, ۱۹۹٦) درسال ۱۹۹۶ اوهلسون و آلبرگ معیارهای پیچیدگی را به منظور پیش‌بینی ماژول‌های از کد به کار بردند که در طول اجرا دچار خطا می‌شدند. مدل پیش گویی کننده آنها ۲۰ درصد از ماژول‌هایی از کد را شناسایی می‌کرد که شامل ۴۷ درصد از کل خطاها بودند.(Ohlsson & Alberg, ۱۹۹٦) زائو در سال ۱۹۹۸ به بررسی معیارهای طراحی نرم‌افزار پرداخت و عملکرد آنها در پیش‌بینی خطا را محاسبه نمود. زائو نشان داد که معیارهای طراحی نیز به خوبی معیارهای کد در پیش‌بینی خطا عمل می‌نمایند. (Zhao, Wohlin, Ohlsson, & Xie, ۱۹۹۸) با پژوهش بر روی ۵ پروژه ماکروسافت[۳۲] در سال ۲۰۰۶، ناپاگان معیارهایی برای پیش‌بینی خطاهای پس از زمان انتشار شناسایی کرد و گزارش مبنی بر اینکه چگونه به طور سیستماتیک پیش‌بینی کننده‌های خطاهای پس از زمان انتشار بر اساس تارخچه[۳۳] ساخته شوند، ارائه داد. (Nagappan, Ball, & Zeller, ۲۰۰٦) شروتر، زیمرمن و زلر در

[۳۲] Microsoft
[۳۳] history