دانشکدگان علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

بهبود روش وارسی مدل با استفاده از تعبیر مجرد

نگارنده: **پویا پرتو**

استاد راهنمای اول: دکتر مجید علی‌زاده
استاد راهنمای دوم: دکتر مجتبی مجتهدی

پایان نامه برای دریافت درجه کارشناسی ارشد
در رشته علوم کامپیوتر

تاریخ دفاع

چکیده

روش وارسی مدل یک روش قابل اعتماد برای بررسی صحت عملکرد برنامه‌های کامپیوتری است. بیان‌های مختلف این روش از منطق موجهات بهره می‌برند که چندان برای برنامه نویسان شناخته شده نیستند. در این رساله سعی شده است که یک بیان جدید از روش وارسی مدل مورد شرح و بررسی قرار گیرد که در ادبیات نظریه تعبیر مجرد بیان شده است و در آن به جای منطق موجهات از عبارات منظم استفاده شده است.

پس از ارائه‌ی مفاهیم اولیه، به سه صورت متفاوت به بیانی جدید از روش وارسی مدل پرداخته‌ایم. صورت اول ساختار خاصی ندارد و صرفا در ادبیات نو بیان شده است، صورت دوم ساختار عبارات منظم را به صورتبندی‌اش اضافه کرده است و در صورت سوم، با اضافه شدن ساختار برنامه به صورتبندی، روش به پیاده سازی نزدیک‌تر شده است. معادل بودن این سه صورت نیز مطالعه و بررسی می‌شود.

کلمات کلیدی: وارسی مدل، نظریه تعبیر مجرد، معناشناسی دلالتی، درستی‌یابی صوری، تحلیل ایستا، درستی‌یابی برنامه‌های کامپیوتری

تقدیم

این پایان نامه را تقدیم می‌کنم خدمت استاد ارجمندم، جناب آقای دکتر امیر عباس ورشوی، به پاس تاثیر عمیقی که بر زندگی من در چند سال گذشته داشته‌اند.

سپاسگزاری

ابتدا، از خانواده‌ی عزیزم سپاسگزاری می‌کنم، بابت اینکه هر چه که توانسته‌اند را در پشتیبانی معنوی و مادی از من دریغ نکرده‌اند.

پس از آن، از دو استاد ارجمندم جناب آقایان دکتر مجید علی‌زاده و دکتر مجتبی مجتهدی بابت زحمات و حمایت‌هایشان سپاسگزاری می‌کنم. بسیار مفتخرم که در این مدت دانشجوی ایشان بوده‌ام.

در نهایت از دوستان عزیزم در داخل و بیرون از دانشگاه سپاسگزاری می‌کنم که طی این مدت در کنار من بودند. به طور ویژه، از جناب آقای علیرضا محمودیان بابت حضور ارزشمند و کمک‌هایشان سپاسگزاری می‌کنم.

پیشگفتار

با توجه به پیشرفت روز افزون علوم کامپیوتر و ورود کاربردهای آن به زندگی روزمره، پیشرفت در روش‌های ساخت و نگه‌داری برنامه‌ها نیازی آشکار به نظر می‌رسد. یکی از مسائل مهم در این زمینه بررسی صحت کارکرد برنامه‌های نوشته‌شده است. عدم صحت کارکرد برنامه‌های نوشته شده بسته به حساسیت یک برنامه می‌تواند تبعات زیان‌بار جبران ناپذیری به همراه داشته‌باشد. پرتاب ناموفق آریان ۵[؟] ، از مدار خارج شدن مدارگرد مریخ [؟] و تصادف هلیکوپتر چینوک [؟] چند نمونه از تبعات بزرگ این قضیه در گذشته بوده اند، همین طور به‌سادگی می‌توان فجایع دیگری از این دست را در زندگی روزمره‌ی انسان‌ها متصور شد.

برای تعیین صحت کارکرد برنامه‌های کامپیوتری روش های متفاوتی ابداع شده‌اند که در ادامه به طور مختصر از آن‌ها یاد می‌کنیم، اما پیش از آن به یک خاصیت مشترک همه‌ی این روش‌ها، یعنی "ناکامل بودن"، می‌پردازیم. منظور از ناکامل بودن این است که با استفاده از هیچ یک از روش‌هایی که داریم، نمی‌توانیم هر خاصیتی را برای هر برنامه‌ای بررسی کنیم. به عبارت دیگر، استفاده از هر روشی محدودیت‌هایی دارد. البته قضیه رایس [؟] به ما این تضمین را داده که روش کاملی اصلا وجود ندارد. قضیه رایس(به طور غیر رسمی) بیان می‌کند که مسئله ی بررسی هر خاصیت غیر بدیهی، برای همه ی برنامه‌ها، تصمیم ناپذیر است. این دلیلی بر این شده که روش‌های مختلفی برای این کار معرفی شوند که هر کدام می‌توانند حالت‌های خاصی از مسئله را حل کنند.

یک دسته‌بندی برای این روش‌ها تقسیم آن‌ها به دو دسته‌ی پویا و ایستا است. روش‌های پویا روش‌هایی هستند که در آن‌ها تست برنامه همزمان با اجرای برنامه است، درحالیکه روش‌های ایستا بدون اجرای برنامه‌ها آن‌ها را تست می‌کنند.

روش‌های پویا معمولاً با اجرای حالات محدودی از برنامه تصمیم می‌گیرند که برنامه‌ای که نوشته شده است، انتظارات را برآورده می‌کند یا خیر. اگر این روش بتواند تشخیص دهد که برنامه‌ای درست کار نمی‌کند، می‌توانیم با اطمینان نتیجه بگیریم که آن برنامه غلط نوشته‌شده است، اما اگر

برنامه‌ای از تست‌های ساخته‌شده با این روش‌ها با موفقیت عبور کند، نمی‌توان اطمینان حاصل کرد که برنامه درست کار می‌کند، زیرا ممکن است، حالتی مشکل‌زا از اجرای برنامه وجود داشته باشد که در تست ها نیامده باشد. برای اطلاعات بیشتر به [؟] مراجعه کنید.

روش‌های ایستا معمولاً روش‌هایی هستند که از نظریه‌های مختلف در منطق ریاضی به عنوان ابزار بهره می‌برند تا بدون اجرای خود برنامه‌ها در مورد صحت اجرای آن‌ها نتیجه‌گیری کنند. به همین دلیل به بخشی مهم و بزرگی از این دستورات که از منطق استفاده می‌کنند روش‌های صوری هم گفته‌می‌شود. معروف ترین روش‌های ایستا؛ روش وارسی مدل، روش‌های استنتاجی و استفاده از نظریه تعبیر مجرد است.

در روش وارسی مدل، یک مدل صوری متناهی از برنامه‌ی موردبررسی می‌سازیم که همه‌ی حالات اجرای برنامه با آن قابل‌توصیف است، سپس با استفاده از یک زبان صوری که بتواند در مورد مدل هایمان صحبت‌کند، ویژگی‌های مورد بررسی را بیان‌می‌کنیم و در نهایت صحت ویژگی‌های بیان‌شده را بررسی می کنیم. مقاله [؟] شروع این روش‌ها بوده که این کار را با استفاده از نوعی مدل کریپکی [؟] و نوعی منطق زمانی به نام منطق زمانی خطی [؟] انجام داده که روشی است با دقت و البته هزینه‌ی محاسباتی بسیار بالا. [؟] یک منبع بسیار مقدماتی و کتاب[؟] یک مرجع سنتی در این زمینه است.

در روش‌های استنتاجی که شاید بتوان یکی از ابتدایی ترین آن‌ها را استفاده از منطق هور[؟] دانست، درستی کارکرد برنامه‌هایمان را با ارائه‌ی یک درخت اثبات در یک دستگاه استنتاجی، متناسب با زبان برنامه‌هایمان، نشان می‌دهیم. در این روش هم اگر بتوانیم درستی یک برنامه را اثبات کنیم، دیگر به طور نظری، خیالی آسوده از درستی برنامه خواهیم داشت، اما ساختن درخت اثبات در یک نظریه برهان می‌تواند چالش برانگیز باشد. در[؟] به منطق هور به طور مقدماتی پرداخته شده است. همین طور کتاب[؟] نیز به پیاده‌سازی منطق هور در زبان coq پرداخته است، که در آن coq یک اثبات‌یار است که بر اساس نظریه نوع وابسته کار می‌کند. برای اطلاعات بیشتر در مورد چگونگی طرز کار این اثبات‌یار و نظریه‌ی بنیادین آن به کتاب[؟] مراجعه کنید. نظریه‌ی مورد شرح در[؟] نیز می‌تواند در این مسیر به کار گرفته شود.

نظریه تعبیر مجرد[؟] نیز یک نظریه ریاضیاتی است که به‌نوعی سعی می‌کند از روی معناشناسی یک برنامه‌ی کامپیوتری[؟] یک تقریب بسازد. منظور از تقریب یک دستگاه کوچک‌تر از معناشناسی اصلی است که رفتارش زیرمجموعه‌ی رفتارهای دستگاه اصلی است. سعی بر این است که دستگاه جدیدی که می‌سازیم به لحاظ محاسباتی ساده‌تر از معناشناسی اصلی کار کند تا بتوان خواص آن را راحت‌تر بررسی کرد. در این صورت هر نتیجه‌ای در مورد خواص جدید، را می‌توان

برای خود برنامه هم بیان کرد، اما توجه شود که در این صورت ممکن است به همه‌ی حقایق دست پیدا نکنیم. برای اطلاعات بیشتر به [؟] و [؟] مراجعه شود.

در این پایان نامه، تمرکز ما روی بهبود روش اول یعنی روش وارسی مدل به کمک روش سوم یعنی نظریه‌ی تعبیر مجرد خواهد بود.

فهرست مطالب

۱۰۰ نظریه تعبیر مجرد

به‌طور خلاصه، نظریه تعبیر مجرد یک چارچوب برای ساختن یک تقریب از معناشناسی یک زبان برنامه نویسی است.

معناشناسی یک زبان یک مدل ریاضیاتی مجرد است که چگونگی رفتار برنامه‌ها در این زبان را توصیف می‌کند. تقریب نیز یک معناشناسی دیگر است که قرار است بخشی(نه همه) از رفتارهای یک برنامه‌ی کامپیوتری در حال اجرا در یک زبان را توصیف کند. این که تقریب چیست، یک معناشناسی را در چه زمانی می‌توانیم تقریبی برای معناشناسی دیگری بدانیم و از یک تقریب چه چیزهایی را می‌توانیم بفهمیم و مواردی دیگر در مورد ارتباط بین دو مدل ریاضیاتی که درباره‌ی معنای برنامه‌ها در یک زبان برنامه‌نویسی واحد صحبت می‌کنند، همگی موضوع بحث در نظریه‌ي تعبیر مجرد است. پس تا اینجا مشخص شد که نظریه‌ی تعبیر مجرد در مورد ارتباط بین معناشناسی‌های مختلف صحبت می‌کند.

برای شروع بحث صوری در مورد این نظریه، از مفهوم دامنه و معناشناسی شروع می‌کنیم. در واقع، این نوع از مشخص کردن معناشناسی یک زبان برنامه نویسی را معناشناسی دِلالَتی نامیده‌اند. در فصول آینده با یک معناشناسی از این نوع سر و کار خواهیم داشت.

تعریف ۱۰۰. (*معناشناسی و دامنه*): *اگر* $\mathbb{P}$ *مجموعه‌ی برنامه‌ها در یک زبان برنامه نویسی باشد، به تابع* $S : \mathbb{P} \to D$ *یک معناشناسی و به مجموعه‌ی* D *یک دامنه می‌گوییم.*

همان‌طور که از تعریف مشخص است، برای این که بتوانیم معنای برنامه‌های کامپیوتری موجود در یک زبان را تعریف کنیم، به یک مجموعه به اسم دامنه احتیاج داریم. تلاش برای پی بردن به این که در یک معناشناسی باید چه مجموعه‌ای را به عنوان دامنه در نظر گرفت، منجر به تولد یک مبحث به نام نظریه‌ی دامنه شده است.

در فصل‌های بعدی، با یک معناشناسی دلالتی سر و کار خواهیم داشت. پس از ارائه‌ی یک زبان برنامه نویسی، یک معناشناسی برای آن زبان معرفی می‌کنیم که معناشناسی رد پیشوندی نام دارد. در این معناشناسی، دامنه یک مجموعه است که شامل موجوداتی به نام رد پیشوندی است. هر رد پیشوندی یک دنباله است که در هر عضو آن اطلاعات موجود در حافظه و مرحله‌ی اجرای برنامه مشخص شده است.

اما فعلا که در حال صحبت در مورد نظریه‌ی تعبیر مجرد هستیم، معناشناسی خاصی را معرفی نمی‌کنیم و بحث را کلی‌تر پیش می‌بریم. نظریه تعبیر مجرد برای معناشناسی‌ها یک چارچوب

مشخص کرده و فقط در مورد معناشناسی‌هایی که در این چارچوب می‌گنجند می‌تواند صحبت کند. یکی از محدودیت‌های این چارچوب این است که دامنه باید یک ترتیب جزئی باشد.

تعریف ۲.۰ (*ترتیب جزئی*): *یک مجموعه‌ی D را به همراه یک رابطه‌ی $\leq$ روی آن مجموعه ترتیب جزئی می‌گوییم، اگر و تنها اگر خواص زیر را داشته باشند:*

$$\blacktriangleleft\ \forall a \in D : a \leq a$$

$$\blacktriangleleft\ \forall a, b \in D : a \leq b \wedge b \leq a \rightarrow a = b$$

$$\blacktriangleleft\ \forall a, b, c \in D : a \leq b \wedge b \leq c \rightarrow a \leq c$$

حال به تعریف بخش بزرگتری از این چارچوب می‌پردازیم. در جبر مجرد مفهومی به اسم تناظر گالوا وجود دارد. این تناظر بین مجموعه‌ای از گروه‌ها و مجموعه‌ای از توسیع میدان‌هایی خاص وجود دارد که به بحث ما مربوط نمی‌شوند. این تناظر یک شکل نظریه ترتیبی هم دارد که در آن به جای مجموعه‌ای از گروه‌ها و میدان‌ها، دو مجموعه‌ی جزئا مرتب داریم. می‌توان گفت در واقع این یک مجرد سازی تناظری است که از جبر آمده.

به شکل ضعیف‌تر نظریه ترتیبی این تناظر اتصال گالوا می‌گویند که در نظریه تعبیر مجرد به عنوان شرط تقریب تعریف شده است، به این معنی که دامنه‌ی یک معناشناسی باید با دامنه‌ی تقریبش یک اتصال گالوا داشته باشد.

تعریف ۳.۰ (*اتصال گالوا*): *برای دو ترتیب جزئی $(A, \leq)$ و $(C, \subseteq)$ زوج $\langle \alpha, \gamma \rangle$ شامل دو تابع $\gamma : A \rightarrow C$ و $\alpha : C \rightarrow A$، یک اتصال گالوا است اگر و تنها اگر*

$$\forall c \in C : \forall a \in A : \alpha(c) \leq a \leftrightarrow c \subseteq \gamma(a)$$

.

نتیجه گیری

دیدیم که صوری‌سازی روش وارسی مدل در ادبیات نظریه‌ی تعبیر مجرد به چه شکل است و در همین حال تلاش کردیم این صوری‌سازی را شفاف‌تر و واضح‌تر از [؟] بیان کنیم. همین طور دیدیم که می‌توان به جای منطق‌های زمانی از عبارات منظم در روش وارسی مدل استفاده کرد.

همان طور که در [؟] آمده است و در فصل دوم به‌طور واضح‌تری نشان دادیم، این روش قابل پیاده‌سازی نیست. در [؟] نزدیک کردن کار به پیاده‌سازی را از طریق متناهی کردن مجموعه‌ی وضعیت‌ها ممکن می‌داند. به هر حال، در واقعیت هم مجموعه‌ی وضعیت‌ها متناهی است، چون حافظه‌ها متناهی هستند و این فرض می‌تواند صوری‌سازی را یک قدم به واقعیت نزدیک‌تر کند.

ایده‌ای که ما برای نزدیک کردن این کار به پیاده‌سازی داریم این است که عبارات منظم محدودتر شود. اگر علاوه‌بر متناهی کردن مجموعه‌ی وضعیت‌ها، دو عملگر * و $^+$ را از عبارات منظم حذف کنیم، صوری‌سازی احتمالا قابل پیاده‌سازی خواهد شد. درست است که حذف این دو عملگر از قدرت بیان ویژگی‌ها کم می‌کند، اما شاید در عمل، همان قدرت بیان باقی مانده برای بیان بسیاری از ویژگی‌ها کافی باشد.

Bibliography

Abstract

Model checking is a reliable method for program verification. Different forms of this method use temporal logic to express properties, which is not commonly accepted by programmers. In this thesis, it has been tried to represent a new form of model checking that has been stated in the literature of abstract interpretation theory and uses regular expressions for expressing program properties instead of modal logic.

After representing basic notions, three novel forms of model checking have been introduced. The first form has no structure and is merely expressed in a new literature; the second form has added the structure of regular expressions to itself; and the third form has the structure of programs in it to get closer to implementation. The equivalence of the three forms has been studied and discussed as well.

Kew Words: Model Checking, Abstract Interpretation, Denotational Semantics, Formal Verification, Static Analysis, Formal Verification of Computer Programs

College of Science
School of Mathematics, Statistics, and Computer Science

Improving Model Checking with Abstract Interpretation

Pouya Partow

Supervisor: Majid Alizadeh
Co-Supervisor: Mojtaba Mojtahedi

A thesis submitted to Graduate Studies Office
in partial fulfillment of the requirements for the degree of
Master of Science in
Computer Science

2023