

Streszczenie

Formalna weryfikacja zdolności strategicznych w systemach wieloagentowych (MAS) jest zadaniem krytycznym, ale niezwykle trudnym, szczególnie w warunkach niedoskonałych informacji i niedoskonałej pamięci. Nieodłączna złożoność obliczeniowa, często charakteryzująca się eksplozją przestrzeni stanów i wysoką złożonością teoretyczną decyzji, sprawia, że dokładne sprawdzanie modeli takich systemów jest niemożliwe w wielu rzeczywistych zastosowaniach. W niniejszej pracy podjęto te fundamentalne wyzwania poprzez opracowanie, wdrożenie i empiryczną ocenę zestawu nowatorskich metod skalowalnej weryfikacji właściwości strategicznych w złożonych środowiskach o ograniczonej dostępności informacji.

Naszym głównym wkładem jest kompleksowa struktura służąca do przybliżonej weryfikacji zdolności strategicznych, sformalizowana przy użyciu logiki temporalnej czasu alternującego z niepełną informacją i strategiami bez pamięci ($\mathbf{ATL}_{ir}$). Wprowadzamy nowatorski schemat aproksymacji stałopunktowej, który zapewnia solidne dolne i górne granice wartości logicznej formuł $\mathbf{ATL}_{ir}$. Osiąga się to poprzez przetłumaczenie specyfikacji $\mathbf{ATL}_{ir}$ na zmienną wersję epistemicznego μ -rachunku, wzbogaconą o operator stałego następnego kroku, który rejestruje zaangażowanie agentów w strategię w ramach ich wspólnej wiedzy. Aby zarządzać kombinatoryczną złożonością relacji epistemicznych, opracowujemy zoptymalizowane struktury danych oparte na lasach rozłącznych zbiorów i proponujemy dynamiczne techniki redukcji przestrzeni stanów, w tym abstrakcję modelu i redukcje częściowo-porządkowe, które można synergicznie połączyć z naszymi metodami punktów stałych. W przypadku systemów wykazujących interakcje asynchroniczne wprowadzamy technikę lokalnego przybliżenia modelu. Dzięki skonstruowaniu abstrakcji globalnego systemu skoncentrowanych na agentach metoda ta umożliwia weryfikację zdolności strategicznych przy

znacznej redukcji przestrzeni stanów, zachowując jednocześnie formalne gwarancje poprawności.

Ponadto przedstawiamy algorytm syntezy strategii oparty na wyszukiwaniu w głąb, wzbogacony o nowatorską relację dominacji dla strategii częściowych. Relacja ta umożliwia systematyczne eliminowanie suboptymalnych wyborów, znacznie usprawniając wyszukiwanie strategii prowadzących do zwycięstwa. Rozszerzamy to podejście na wielokryterialną optymalizację strategii, umożliwiając syntezę strategii, które są jednocześnie skuteczne w odniesieniu do wielu celów, takich jak ograniczenie wyników i jednolitość działań.

Aby zweryfikować praktyczną skuteczność naszych teoretycznych osiągnięć, opracowaliśmy StraTegic Verifier (STV), najnowocześniejszy weryfikator modelowy, który implementuje proponowane algorytmy. Zastosowaliśmy STV i nasze metody do szerokiej gamy trudnych modeli rzeczywistych, w tym gry karcianej brydż, autonomicznych zespołów dronów, inteligentnych fabryk produkcyjnych, ekosystemów społecznie wyjaśnialnych sztucznych inteligencji (SAI) oraz zaawansowanych protokołów głosowania elektronicznego, takich jak Selene. Wyniki naszych eksperymentów pokazują, że proponowane metody aproksymacji i syntezy zapewniają znaczną poprawę wydajności, często o kilka rzędów wielkości, w porównaniu z istniejącymi najnowocześniejszymi narzędziami. W wielu przypadkach nasze podejście pozwoliło z powodzeniem zweryfikować modele zawierające miliony stanów, co jest niemożliwe do osiągnięcia przy użyciu metod dokładnych, przy zachowaniu wysokiego stopnia dokładności i zapewnieniu jednoznacznych wyników.

Podsumowując, niniejsza praca ustanawia solidną i skalowalną metodologię weryfikacji zdolności strategicznych w warunkach niedoskonałej informacji. Wypełniając lukę między logiką teoretyczną a praktyczną implementacją, nasza praca umożliwia formalną analizę złożonych systemów wieloagentowych, które wcześniej były poza zasięgiem automatycznej weryfikacji, przyczyniając się tym samym do rozwoju zarówno teorii, jak i praktyki zapewniania poprawności i bezpieczeństwa w coraz bardziej autonomicznych i interaktywnych środowiskach komputerowych.

Abstract

The formal verification of strategic abilities in multi-agent systems (MAS) is a critical yet notoriously challenging endeavor, particularly under conditions of imperfect information and imperfect recall. The inherent computational complexity, often characterized by state-space explosion and high decision-theoretic complexity, renders the exact model checking of such systems intractable for many real-world applications. This thesis addresses these fundamental challenges by developing, implementing, and empirically evaluating a suite of novel methods for the scalable verification of strategic properties in complex, informationally restricted environments.

Our primary contribution is a comprehensive framework for approximate verification of strategic ability, formalized using Alternating-time Temporal Logic with imperfect information and memoryless strategies (**ATL_{ir}**). We introduce a novel fixpoint approximation scheme that provides sound lower and upper bounds for the truth value of **ATL_{ir}** formulae. This is achieved by translating **ATL_{ir}** specifications into an alternating epistemic μ -calculus variant, enhanced with a steadfast next-step operator that captures agents' commitment to strategies within their common knowledge neighborhoods. To manage the combinatorial complexity of epistemic relations, we devise optimized data structures based on disjoint-set forests and propose dynamic state-space reduction techniques, including model abstraction and partial-order reduction, which can be synergistically combined with our fixpoint methods. For systems exhibiting asynchronous interaction, we introduce a local model approximation technique. By constructing agent-centric abstractions of the global system, this method enables verification of strategic abilities with a substantial reduction in the state space, while preserving formal soundness guarantees.

Furthermore, we present a forward-chaining strategy synthesis algorithm based on depth-first search, enhanced with a novel dominance relation for

partial strategies. This relation enables the systematic pruning of suboptimal choices, significantly improving the search for winning strategies. We extend this approach to multi-criteria strategy optimization, allowing for the synthesis of strategies that are simultaneously effective with respect to multiple objectives, such as outcome containment and action uniformity.

To validate the practical efficacy of our theoretical contributions, we developed the *StraTegic Verifier* (STV), a state-of-the-art model checker that implements the proposed algorithms. We applied STV and our methods to a diverse range of challenging real-world models, including the bridge card game, autonomous drone teams, smart production factories, social explainable AI (SAI) ecosystems, and sophisticated e-voting protocols like Selene. Our experimental results demonstrate that the proposed approximation and synthesis methods yield dramatic performance improvements, often by several orders of magnitude, over existing state-of-the-art tools. In numerous instances, our approach successfully verified models with millions of states, a feat unattainable by exact methods, while maintaining a high degree of accuracy and providing conclusive results.

In conclusion, this thesis establishes a robust and scalable methodology for the verification of strategic abilities under imperfect information. By bridging the gap between theoretical logic and practical implementation, our work enables the formal analysis of complex multi-agent systems that were previously beyond the reach of automated verification, thereby advancing both the theory and practice of ensuring correctness and security in increasingly autonomous and interactive computing environments.