دانشکده مهندسی برق و کامپیوتر

پردیس دانشکده های فنی

بسمه تعالی

جلســه دفاعیه پایان نامه کارشناسی ارشد

گرایش: هوش مصنوعی و رباتیکز

موضوع: یادگیری تقویتی چندعامله در بازی های مارکوف با دیدگاه نظریه چشم انداز تجمعی

توسط: حافظ قائمی

استاد راهنما: دکتر حامد کبریایی – دکتر مجید نیلی احمدآبادی

استاد مشاور: -

روز ، ساعت ، تاریخ دفاع : چهارشنبه ۱۸ مرداد ۱۴۰۲، ساعت ۸:۳۰ تا ۱۰:۰۰

مکان دفاع : اتاق ۸۱۴، ساختمان شماره ۲ دانشکده مهندسی برق و کامپیوتر، پردیس فنی دانشگاه تهران

چکیده:

درونزادی و حساسیت به ریسک در یادگیری تقویتی، در محیط‌های تک‌عامله یا چندعامله، چارچوبی است که در مقایسه با ناسوی‌داری کامل عامل‌ها کمتر به آن پرداخته شده است. یک معیار ریسک را می‌توان یا با دگرگونی تابع هدف (ریسک صریح) یا با افزودن قید (ریسک ضمنی) به مسئله بهینه‌سازی عامل(ها) وارد یادگیری تقویتی کرد. هدف این پایان‌نامه توسعه یک چارچوب برخط بدون مدل برای یادگیری تقویتی حساس به مخاطره در یادگیری تقویتی تک‌عامله و یادگیری تقویتی چندعامله در بازی‌های مارکوف است. به این منظور، به عنوان معیار ریسک، از نظریه چشم‌انداز تجمعی که معیار ریسکی صریح و غیر محدب است استفاده می‌کنیم. نظریه چشم‌انداز تجمعی مدلی در اقتصاد عصبی است که می‌تواند زیان‌گریزی انسان‌ها و گرایش آنها به فزون برآورد احتمالات پایین و کاست برآورد احتمالات بالا را توضیح دهد. برای توسعه چارچوب گفته شده، دو الگوریتم عملگر-نقاد برپایه نمونه‌گیری و حساس به معیار ریسک نظریه چشم‌انداز تجمعی ارائه می‌کنیم. برای الگوریتم اول که در قالب یادگیری تقویتی تک‌عامله ارائه شده‌است، اثبات می‌کنیم که سیاست عامل به صورت مجانبی به سیاست بهینه محلی درونزاد همگرا می‌شود. این همگرایی را به صورت تجربی نیز با انجام آزمایش‌هایی در محیطی که برای مطالعه استراتژی‌های قیمت‌گذاری ایستگاه‌های سوخت‌گیری خودروهای برقی توسعه داده شده‌است نشان می‌دهیم. نتایج نشان می‌دهند سیاست‌های به دست‌آمده توسط الگوریتم حساس به ریسک پیشنهادشده در این آزمایش‌ها نسبت به سیاست‌های به دست‌آمده توسط الگوریتم عملگر-نقاد غیرحساس به ریسک عملکرد بهتری در بهینه‌سازی تابع ارزش درونزاد عامل دارند. برای الگوریتم دوم که در قالب یادگیری تقویتی چندعامله غیرمتمرکز در بازی‌های شبکه‌ای تجمعی مارکوف ارائه شده‌است، تحت شرایط معقولی اثبات می‌کنیم که سیاست عامل‌ها به صورت مجانبی به مفهوم درونزادی از تعادل کامل نش مارکوف همگرا می‌شود. نتایج تجربی همگرایی این الگوریتم نشان داده می‌دهند سیاست‌های نش درونزاد برمبنای نظریه چشم‌انداز تجمعی می‌توانند متفاوت با سیاست‌های نش بدون ریسک باشند و این تفاوت با توجه به میزان زیان‌گریزی عامل قابل تفسیر است. چارچوب توسعه داده‌شده در این پایان‌نامه اولین خانواده الگوریتم‌های بدون مدل در یادگیری تقویتی حساس به ریسک با معیار ریسک نظریه چشم‌انداز تجمعی است.