



УНИВЕРЗИТЕТ У БЕОГРАДУ - ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

Булевар краља Александра 73, 11000 Београд, Србија

Тел. 011/324-8464, Факс: 011/324-8681

КОМИСИЈИ ЗА СТУДИЈЕ II СТЕПЕНА ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА У БЕОГРАДУ

Комисија за студије II степена, Електротехничког факултета у Београду, на својој седници одржаној 28.08.2018. године именовало нас је у Комисију за преглед и оцену мастер рада дипл. инж. Наталије Ђурић под насловом „Алгоритам дубоког детерминистичког градијента управљања за балансирање инверзним клатном“. Након прегледа материјала Комисија подноси следећи

ИЗВЕШТАЈ

1. Биографски подаци кандидата

Наталија Ђурић је рођена 17.09.1994. године у Прокупљу. Завршила је основну школу "Ратко Павловић Ђићко" у Прокупљу као вуковац. Уписала је Гимназију у Прокупљу коју је такође завршила као вуковац. Електротехнички факултет уписала је 2013. године. Дипломирала је на одсеку за Сигнале и системе 2017. године са просечном оценом на испитима 8.93, на дипломском 10. Дипломске академске – мастер студије на Електротехничком факултету у Београду, на Модулу за сигнале и системе уписала је у октобру 2017. године. Положила је све испите са просечном оценом 9.8.

2. Опис мастер рада

Мастер рад обухвата 47 страна, са укупно 22 слике, 4 табеле и 14 референци. Рад садржи 5 поглавља (укључујући закључак), прилог са написаним кодом и списак коришћене литературе.

Прво поглавље представља увод у коме је укратко описан појам дубоког учења подстицањем, уз осврт на данас доминантне приступе у овој области: методе засноване на 1) вредносу функцији, 2) градијенту управљања и 3) методе које комбинују оба приступа (тзв. *actor-critic*, односно актант-критичар методе).

У другом поглављу су описани теоријски основи алгоритма. Дефинисан је проблем учења подстицањем насупрот стандардном машинском учењу са надзором и учењу без надзора. Описан је Марковљев процес одлучивања (МПО) као основни математички модел у проблемима учења подстицањем. Дефинисани су појмови функција вредности стања и вредности акција. Након тога су описане неуралне мреже, које се у дубоком учењу подстицањем користе као апроксиматори функције вредности акције, и за пресликовање стања у одговарајућу акцију.

Треће поглавље садржи опис алгоритма градијента детерминистичког управљања (енгл. *deep deterministic policy gradient*, DDPG), данас једном од најпопуларнијих приступа учењу подстицањем у проблемима са континуалним простором акција. Објашњена је архитектура овог алгоритма, чију суштину чине две мреже: једна којом се стања пресликају у акције (актант) и друга којом се стања и акције пресликају у тзв. Q-вредности (критичар). Указано је на проблеме корелисаности примера у обуčавајућем скупу и нестабилности која је узрокована чињеницом да се циљни излази ових мрежа рачунају на основу њихових тренутних, стварних излаза, што доводи до проблема са стабилношћу обуčавања. Описаны су стандардни приступи превазилажењу ових проблема: понављање искуства и увођење циљних мрежа. Објашњена је потреба за истраживањем окружења, што се остварује додавањем случајног шума на акције које предлаже актант.

У четвртом поглављу је описано OpenAI Gym окружење које је коришћено за верификацију имплементираног алгоритма. Конкретно, коришћен је симулатор инверзног клатна, које представља један од класичних проблема у теорији аутоматског управљања. Приказан је ток обучавања алгоритма и показано је да долази до конвергенције ка стационарном закону управљања. Анализирано је понашање обученог агента у неколико карактеристичних ситуација. Поређењем са подацима из литературе, утврђено је да су резултати остварени у оквиру овог рада приближно једнаки најбољима који су до сада остварени на овом конкретном тест примеру.

У закључку је кратко рекапитулиран садржај рада и указано на неке од добрих особина анализираног алгоритма.

3. Анализа рада са кључним резултатима

Мастер рад дипл. инж. Наталије Ђурић се бави проблемом дубоког учења подстицањем. Конкретно, у раду је описан и имплементиран алгоритам заснован на градијенту детерминистичког закона управљања, у литератури познат под називом DDPG. Алгоритам је детаљно теоријски описан, имплементиран у језику Python 3.6 коришћењем библиотеке TensorFlow, и верификован применом на симулатору инверзног клатна доступном у оквиру библиотеке OpenAI Gym.

Кандидаткиња је темељно прегледала релевантну савремену стручну литературу и детаљно теоријски описала све компоненте DDPG алгоритма. Након тога је прецизно обrazложила начин имплементације алгоритма, објаснила на који начин је извршила верификацију и критички анализирала остварене резултате.

Основни доприноси рада су: 1) детаљан преглед и критичка анализа литературе из области дубоког учења са подстицањем, са акцентом на DDPG алгоритму; 2) имплементација DDPG алгоритма и његова верификација на једном од стандардних проблема аутоматског управљања са континуалним простором акција; 3) критички су анализирани остварени резултати и особине примењеног алгоритма.

4. Закључак и предлог

Кандидаткиња Наталија Ђурић је у свом мастер раду успешно применила алгоритам дубоког градијента детерминистичког управљања у циљу имплементације агента који самостално учи да балансира инверзно клатно. При томе је исказала самосталност, систематичност и иновативност у раду, као и способност коришћења релевантне литературе.

На основу изложеног, Комисија предлаже Комисији за студије II степена Електротехничког факултета у Београду да рад дипл. инж. Наталије Ђурић прихвати као мастер рад и кандидаткињи одобри јавну усмену одбрану.

Београд, 10.09.2018. године

Чланови комисије:

П. Јагић

др Предраг Тадић, доцент

др Горан Квашчев, ванредни професор