



ДЪЛБОКО СТРУКТУРИРАНО ОБУЧЕНИЕ И НЕВРОБИОЛОГИЯТА

1. Симбиотична връзка на дълбоко структурираното обучение и невробиологията.

Изкуственият интелект и изкуствените невронни мрежи са мотивирани и вдъхновени по модел и организация на естествения интелект и биологичните невронни мрежи и от начина, по който се смята, че те работят. Това е отразено в структурното построяване (дизайна) на изкуствените невронни мрежи, които се опитват да имитират сложността на връзките между мозъчните клетки, посредством математическа парадигма и моделиране на информационните процеси в невронната система. Различните области на изследване на изкуствения интелект, като машинно обучение, общуване, действия, възприятия, представяне на знания, решаване на проблеми, автоматизирани разсъждения и планиране, имитират една съществена част от дейностите на човешкия интелект. При решаване на задачи и при обсъждане на въпроси, от областта на изкуствения интелект, независимо от сферите на изследване, получените резултати винаги се сравняват и се прави аналогия с начина, по който човек би подходил към решаването на конкретна задача, или би реагирал при взаимодействие с процесите от реалния свят.

Човешкият мозък може да се разглежда като изключително сложна система за обработка на информация, способна да придобива знания, да взема сложни решения, да се учи от опит и да се адаптира успешно към различни предизвикателства и условия на заобикалящия го свят. Той съответства на сложен, нелинеен и паралелен компютър, който притежава потенциалната възможност да възприема до 4 милиарда бита в секунда информация и в същото време да бъде изключително ефективен по отношение на енергията, която използва за своята работа. Редица автори определят човешкия мозък като „усъвършенстван, високопроизводителен биокомпютър, който обработва множество сложни задачи паралелно с висока ефективност и забележително ниска консумация на енергия“. Човешкият мозък е енергийно ефективен и при



своята работа консумира около 30 вата енергия и изразходва около 10-16 джаула за операция на секунда - хиляди пъти по-малко от малък суперкомпютър.

Въпреки интензивните изследвания, които се провеждат, теоретичната невронаука (неврофизиологията) все още не може да разгадае цялостното функциониране и организация на биологичния мозък. Това включва най-общо връзката между структурата и функцията на мозъка. Връзката между различните взаимосвързани нива на протичащи биохимични и физиологични процеси и механизми в мозъчните региони, и причинно-следственото разбиране за начина, по който се случва човешкото познание, поведение и функционалната променливост при отделните индивиди. Знанието, за това как функционира мозъкът, е силно фрагментирано и днес се знае много за отделните му йерархични нива на организация, т. нар. пространствени мащаби, обхващащи девет размерна скала: започваща от гени, протеини (10⁻⁹ метра (м)), хромозоми, синапси, клетки (10⁻⁶ м), до микросхеми, мозъчни региони (10⁻³ м), целия мозък (10⁻² м) и човешкото тяло (100 м).

Все още остава енигма разбирането на мозъка в цялата му сложност – по какъв начин колективните дейности на невроните и моделите им на задействане контролират отделни мозъчни функции, формират паметта и съзнанието. Затова последните години се прилага интердисциплинарен изследователски подход, който използва изчислителната мощност на суперкомпютрите и напредналите методи на изкуствения интелект, невроинформатиката и медицината, в комбинация с дълбоко структурирано обучение, за извършване на новаторско дигитално изследване на мозъка. Това е възможно благодарение на изграждането на споделена научноизследователска инфраструктура за цифрови мозъчни изследвания - Human Brain Project 1 , EBRAINS 2 и Fenix – research infrastructure 3 , финансирана и координирана от Европейския съюз. Десетгодишният проект е базиран на производителността на суперкомпютри и подробно са описани необходимостта от следващо ниво на изчислителна мощност - в мащаб екса (10¹⁸), за да се справят със сложността на човешкия мозък и да успеят да го анализират в цялост, с много висока разделителна способност, на ниво клетъчна структура. Това се доказва и от проведено изследване върху фрагмент от човешка хирургическа проба от мозъчната кора с размер ~ 1 мм³. След изобразяване с високоскоростна електронна микроскопия и направена

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



последваща тримерна реконструкция на сегментираните клетки, клетъчни части, кръвоносни съдове, миелини, инхибиторни и възбуждащи синапси, общият обем от данни достига 1.4 петабайта. Постигнатите високи резултати в отделните направления на изкуствения интелект, напредналите методи на невроинформатиката, както и високия технологичен напредък на компютърната техника, подпомагат широкомащабни и задълбочени изследвания в областта на неврофизиологията.

През последните години дълбокото обучение се справя успешно с разнообразни задачи, като се учи от собствен опит, без да е необходимо изследователите да имат пряка намеса в данните. Такива системи са базирани на неконтролираното обучение - самоконтролирано обучение и механизъм на самовнимание. Предложен са подходи за системно развитие на невронаучните изследвания в ерата на дълбокото обучение. Обсъждат се още динамиката на обучението и невронното представяне в изкуствените и биологичните системи.

Задълбоченото познание на отделни мозъчни механизми на умствени дейности, както и детайлното изучаване на морфологията и начина на действие на нервната система на човек, с всички нейни аспекти – клетъчни, молекулярни, структурни, функционални, еволюционни, изчислителни и др., са вдъхновение за изкуствения интелект. Невроморфното („мозъчно-вдъхновено“) инженерство и изчисления предизвикват нарастващ интерес към основния хардуер, на който работи изкуствения интелект. Започват да се създават изкуствени неврони и синапси, базирани на биологично вдъхновената архитектура на невроните, целта на което да се преодолеят негативните ефекти на машинното обучение – използването на голямо количество енергия, време и данни, необходими за обучението на все по-сложните модели на невронни мрежи. Създават се специализирани „AI чипове“ с имплементирани асинхронни пикови невронни мрежи (Spiking Neural Networks – SNNs) за изпълняване на сложни AI задачи, преосмисляйки фундаментално компютърната архитектура, базирана на транзисторно ниво. Новата област на изкуствения интелект, наречена аналогово дълбоко обучение, се основава на използването на програмируеми резистори (аналогови резисторни превключващи процесори) и мемристори, с които могат да се създават физически невронни мрежи от аналогови изкуствени неврони и синапси, които са биологично правдоподобни невронни модели.

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



Създаването на неорганичен протонен електрохимичен синапс, дава надежда за създаване на нискоенергийни и надеждни изкуствени синапси използвани като хардуер за изкуствени невронни мрежи. Конструирването на синаптични устройства, с консумация на енергия по-ниска от тази на биологичен синапс, са в състояние да вдъхновят напредъка на невроморфните изчисления и да ускорят развитието на изкуствения интелект.

2. Символизъм срещу конекционизъм – дълбоко обучение базирано на конекционистки подход със символни представяния.

Когнитивната наука, невробиологията и изкуствения интелект имат богата споделена история още от началото на миналия век. От Дартмутската конференция, проведена през 1956 г., се обособяват два подхода на развитие на ИИ – конекционизъм и символизъм, които са насочени към интелигентните компютърни системи и използването на различни методи за решаване на поставените задачи. Двата метода са базирани на обработката на знания и информация, и тяхното представяне в рамките на интелигентна компютърна система. Дебатът „символизъм/конекционизъм“ се корени в противопоставяне на методите на двата подхода – знаков (символен) срещу поведенчески (числов). През годините са изтъкнати много аргументи за и против всеки от тях, като диспутът е многоаспектен и се провежда на различни нива – от абстрактното до конкретното.

Конекционисткият (числов, поведенчески) подход е вдъхновен от физиологията на мозъка и се опитва да обясни интелектуалните човешки способности с помощта на изкуствени невронни мрежи, генетични алгоритми и симулационни системи. Основава се на теорията на динамичните системи и се прави опит за симулиране на дейността на централната нервна система, при която начинът на свързване между невроните и предаването на информацията между тях, са носителите на знанието.

Символно базираният (знаков, логически) подход използва предимно формалната логика и се обучава чрез структурирани символни представяния за света, по начин, по който се обучава човешкия ум. Символният модел е насочен към интелигентност, основана на знания, разсъждения, изводи и планиране, с



практически успехи в области като експертни системи, машинен превод, целенасочено планиране, и др.

Последните няколко години се наблюдава възраждане на конекционистката теория и на подхода за паралелна разпределителна обработка (Parallel Distributed Processing - PDP), добили широка популярност през 80-те и 90-те години на миналия век, особено след публикуване на книгата на Дейвид Румелхарт и Джеймс Макклеланд „Parallel Distributed Processing Explorations in the Microstructure of Cognition: Foundation“ („Изследване на паралелна разпределителна обработка в микроструктурата на познанието“, 1986).

Всеки от двата подхода има своите успехи, но е достигнал граници, които не могат да бъдат преодоляни. Диспутът за това кой от двата подхода е по-уместен продължава да занимава академичните среди и до днес. „Неоконекционизъм?“, или „Невро- символичен изкуствен интелект“? Кое е следващото ниво? На ново опит за моделиране на биологични функции, с дълбоки структурирани изчисления, които са способни да решават проблеми от реалния свят, благодарение на символни представяния, големите данни и високата технологична производителност. Румелхарт и МакКлеланд, заявяват в своя труд, че не е задължително PDP моделите да имат т. нар. невронен или биологически реализъм, който може да включва различни степени на биологична правдоподобност, и че ключът към научния прогрес е по-скоро в правилните приближения и правилните опростявания. Въпреки това, много изследователи, подкрепящи конекционисткия подход, се опитват тясно да моделират изкуствените невронни мрежи по биологични аспекти на естествените такива.

Моделите за машинно и дълбоко обучение, през последните няколко години могат да се възползват все повече от символни представяния и системи за разсъждения (формални, символни, обобщение и абстракция (generalization and abstraction) и др.), обединявайки ги чрез прилагане на векторни символни архитектури. Пример за това са CNN архитектури, използвани за кодиране на сложни входни изображения във вектори, за генерирането на високомерни (high-dimensional – HD) векторни представяния. Освен изображения, могат да бъдат кодирани текст и реч. Текстовите данни могат да бъдат представени чрез вектори за задачи като обработка на естествен език. Един вектор може да съставлява буквен знак, дума, абзац или цял документ. По този начин може да се представя



символна информация посредством вектори, вградени във високомерни векторни пространства, като техниките за самоконтролирано обучение – word2vec и GloVe.

Техниката на векторно квантуване (Vector Quantization - VQ) е класически метод за апроксимация и намира приложение в компресиране и кодиране на изображения и реч. Входното пространство се разглежда като поредица от тестови векторни извадки (samples), разделени на региони за вземане на решения, където всеки регион има представителен вектор (кодова дума), а всички кодови думи се обединяват в „кодова книга“. По този начин се изгражда речник на кодовата книга с краен набор от вектори. По този начин цялата съвкупност от данни е сегментирана на региони (области), а всяка кодова дума се намира в свой собствен регион и е представена с даден входен вектор или клъстер от вектори. Техниката VQ се използва и при LVQ невронни мрежи, както и при невронни мрежи- трансформатори като VQGAN + CLIP (Vector Quantized Generative Adversarial Network + Contrastive Language-Image Pre-training). Подобни алгоритми се използват за създаване на произведения на изкуството с помощта на изкуствен интелект на принципа „от текст към изображение“. Например, чрез VQGAN комбинира конволюционни невронни мрежи и трансформатори невронни мрежи, обединявайки ги в генеративна съревнователна мрежа, т. нар. хибриден трансформатор за създаване на художествени изображения. Генеративната съревнователна мрежа противопоставя две невронни мрежи една, срещу друга, като едната има за цел да генерира реалистични изображения, а другата да прави разлика между реални и генерирани изображения. CLIP (предварително обучение на изображение с контрастивен (съпоставителен) език) невронна мрежа, която е от тип самоконтролирано обучение, насочва и управлява VQGAN алгоритъма за това как изображението да прилича повече на въведения текст или фраза, т. нар. текстова подкана.

На фигура 1 са показани изображения, генерирани с помощта на изкуствен интелект и алгоритмите VQGAN + CLIP, с посочени текстови подкани за всяко изображение. Използван е NightCafe Studio Creator за създаването на изображенията показани на фигура 1 и т. нар. Artistic Algorithm (VQGAN + CLIP).



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ
СОЦИАЛЕН ФОНД



ОПЕРАТИВНА ПРОГРАМА
НАУКА И ОБРАЗОВАНИЕ ЗА
ИНТЕЛИГЕНТЕН РАСТЕЖ

Избрани са различни стилове на изобразяване, посочени под текстовата подкана на всяко изображение.



“Artificial Intelligence”
Preset Style: Gouache

(a)



“Artificial Intelligence and
Neural network”
Preset Style: Cubist

(б)



“Black and white image for aircraft
with additive impulse noise”
Preset Style: Charcoal

(в)

Фиг. 1. Създадени изображения с помощта на изкуствен интелект и използвани различни предварително зададени стилове: гваш (a), кубист (б) и въглен (в).

3. Емоционалната машина – въображение и абстракция.

Човешкият мозък на възрастен съдържа около 100 млрд. неврона, свързани с около 100 трилиона връзки, като силата му зависи не от големината, а от сложността на връзките между изграждащите го елементи. Това е една от причините да е толкова мощен, като може да се направи аналогия с компютър, показващ огромен паралелизъм.

Разглеждането на биологичния мозък като масивна паралелна изчислителна машина има своите последователи от 70-те години на миналия век и се свързва с разработването на теорията на Марвин Мински „Society of Mind“ („Общество на разума“), публикувана през 1986 г. В нея човешкият ум се разглежда като огромно общество, съставено от индивидуални прости процеси, наречени агенти, които са „много малки части, всяка безсмислена сама по себе си“, като с помощта на това построение се прави опит да се обясни как интелигентността може да се появи от неинтелигентността.

Всички умствени агенти са кръстосано свързани в общества, в заплетени мрежи чрез объркани пътища. Тази сложна схема изгражда интелекта и съставя

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



неговите способности, откъдето и произтича голяма част от силата на разума, по аналогия на човешкия ум.

Машина и въображение, на пръв поглед две логически изключващи се понятия. Както казва Мински: „Използваме „въображението“ си, за да решим геометричен проблем, да планираме разходка до някое познато място или да изберем какво да ядем за вечеря: във всяко трябва да си представим неща, които всъщност ги няма. Използването на фантазии, емоционални или не, е необходимо за всеки сложен процес на решаване на проблеми. Винаги трябва да се справяме с несъществуващи сцени...“.

През годините Марвин Мински надгражда своите идеи в следваща своя книга „The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind“ („Емоционалната машина: Разумно мислене, изкуствен интелект и бъдещето на човешкия ум“), издадена през 2006 г. В нея той анализира много умствени дейности, разглежда по-широк кръг от психологически теми, разглежда емоциите като различни начини на мислене, които умът ни използва. Заменя „агент“ с „ресурс“, и дава нова обосновка за мозъка като „поддържащ „облак от ресурси“, които взаимодействат по различни начини за да произвеждат различни умствени феномени“. Показва чрез привеждане на примери многостранното човешко мислене, като същевременно непрекъснато прави паралели с възможностите на машината, уподобявайки способностите ѝ с тези на човешките и обратното. Според Мински „въпросът не е дали интелигентните машини могат да имат някакви емоции, а дали машините могат да бъдат интелигентни без каквито и да е емоции“. Машината може да разширява собствения си технологичен обхват на действие, а ограниченията, които според други автори са присъщи на машините (липса на: креативност, гениалност, въображение, изобретателност, интуиция, находчивост, емоционалност и др.), произтичат от собствените ни човешки ограничения.

Според Д. Цанкова, така както емоциите и разума са неразривно свързани части на човешкия интелект, така и изкуствените емоции трябва да заемат полагащото им се място в изкуствения интелект. Изкуствените емоционални модели, като правдоподобните агенти, афективни носещи се компютри, афективни изчисления и емоционалната роботика, включват различни модели на



емоцията, които или употребяват техники от света на изкуствения интелект, или прилагат емоции към различни области на изкуствения интелект.

В противовес на възгледите на Мински, Хюберт Драйфус, отявлен критик на изследванията на изкуствения интелект, разглежда в поредица от свои трудове, философските последици от изкуствения интелект. Първият документ, с който критично разглежда твърдения и предположения от областта е “Alchemy and AI (1965)” (Алхимия и изкуствен интелект). В него Драйфус обобщава историята на изкуствения интелект и атакува оптимизма от постигнатия напредък, който завладява научната област. Неговата критика е насочена главно към Хърбърт Саймън, който твърди и предрича през 1965 г., че „машините ще бъдат способни, в рамките до 20 години, да изпълняват всяка работа, която може да изпълнява човек“.

В следваща своя книга „What Computer Can't Do. A Critique of Artificial Reason (1972)” (Какво компютрите не могат да направят. Критика на изкуствения разум) Драйфус идентифицира и опровергава четири философски допускания, на които се основават ранните изследвания, ентузиазма и вярата, завладели изследователите в сферата. Според него, приеманите от научните среди предположения, като аксиоми и непременно гарантиращи резултати, са погрешни, и опровергава формулираната от Алън Нюел и Хърбърт Саймън хипотеза за физическа символна система.

В своята статия „GPS, A Program that Simulates Human Thought (1961)” (Общо решаване на проблеми, програма, която симулира човешката мисъл) Нюел и Саймън се опитват да покажат, как с помощта на програма за решаване на проблеми като GPS (General Problem-Solving), могат да се конструират теории за човешкото мислене, а поведението на разумния, интелигентен човек, може да се разбира като продукт на сложен, но краен набор от закони.

Според тях манипулирането на символи е в същността както на човешкия, така и на машинния интелект. В своята книга “Mind over Machine (1986)” („Разум над машината“) Хюбърт и Стюърт Драйфус представят аргументи за силата на човешката интуиция пред техническите предимства на машината, като подчертават ограниченията и трудността на изкуствения интелект да се справи със здравия разум. В статията „Why Expert Systems Do Not Exhibit Expertise, (1986)“, (Защо експертните системи не проявяват експертиза), чиито текст е



извлечен от книгата „Mind over Machine...”, братята Драйфус казват: „машинният интелект, следващ правила и манипулиращ символи, вероятно никога няма да замени човешкия интелект, просто защото ние не сме „мислещи машини“. Човешките същества притежават интуитивен интелект, с който „разсъждаващите“ машини просто не могат да се сравнят“.

Пионерът в областта на компютърните науки Лотфи Заде, пише в своя рецензия по повод статията и книгата на братята Драйфус: „това, което най-много дразни общността на изкуствения интелект, е едно от централните твърдения на тази книга, а именно, че машините никога няма да се доближат до хората при изпълнение на когнитивни задачи, изискващи интуиция и холистично мислене“.

Абстрактното описание и представяне на процеси и/или обекти, в компютърните науки и в частност в изкуствения интелект, е механизъм, с който една сложна концепция се свежда до прости инструменти. Абстракцията помага за решаването на даден проблем или за постигане на определена цел, чрез построяването на конкретен път от правила, т. нар. „верига на разсъждения“.

Някои автори дефинират абстракцията като „картографиране (преобразуване) между формализми, които намаляват изчислителната сложност на заложената задача“, в контекста на следните области на изкуствения интелект: доказване на теореми, решаване на проблеми, представяне на знания (пространствени и времеви разсъждения) и машинно обучение. Това може да бъде постигнато чрез използването на различни нива на детайлност (сложност) и способността да се преминава от едно ниво в друго, като същевременно се запазват полезни свойства, оставяйки по-сложните свойства под това ниво. Ключов въпрос остава намирането и формулирането на най- подходящи указания, описващи задачата, което е свързано с изобразяването на различни степени на детайлност. Твърде конкретното и много подробно описание на правила, рядко ще съответства на нови ситуации, докато твърде абстрактното (по-прости описания, потискащи редица подробности), ще съответства на твърде много нови ситуации, което може да отдалечи системата от желания изход.

Изследователят (експертът), с помощта на своя натрупан опит, умения (експертиза), задава пътя на системата, намирайки най- добрата абстракция за



решаване на проблема, като тук е ролята на т. нар. интуиция, находчивост, възприятие.

„Мислене чрез образи, а не думи“ казва Драйфус, още когато компютрите трудно са разпознавали изображения или са анализирали обекти и сцени. Ако Драйфус е прав, че мисловният процес при човека е основан на формиране на образи и тяхното цялостно сравняване, и хората използват изображения, а не символни описания, за да разберат дадена ситуация и да реагират на нея, или за да предскажат определени събития, то тогава невро-символичния подход не е достатъчен за създаването на „мислеща машина“.

Следвайки философските разсъждения на Драйфус, компютърното зрение, може да се окаже ключът или пътят за постигане на следващо ниво изкуствен интелект – от „тесен AI“ и „генеративен AI“ (т.е. творчески – създаващ компютърен код, музика, поезия, изкуство), към постигане на изкуствен общ интелект.

Това може да бъде реализирано заедно с квантовите и невроморфните изчисления и разработените невроморфни изследователски чипове за изкуствен интелект.