

УДК 519.62

Г.Г. Четвериков

АЛГЕБРО-ЛОГІЧНІ ЗАСОБИ МОДЕЛЮВАННЯ ПРИРОДНОЇ МОВИ (ПОВІДОМЛЕННЯ 1)

1. Принципи роботи мозку як відправна точка побудови k -значних структур мовних систем штучного інтелекту

1.1. Вступ

Мислителі цивілізації чітко і ясно окреслили найголовнішу проблему і найкоротший шлях до створення систем штучного інтелекту (ШтІ) – моделювання морально-етичних норм і законів. Очевидно, що єдиним, відомим нам, об'єктивним носієм моралі і інтелекту є людина, а виразником, засобом до зовнішнього спілкування, є людська мова. Отже єдино можливим шляхом найоб'єктивнішого, найповнішого та найефективнішого створення систем ШтІ є шлях аналізу, моделювання та синтезу мовного інтелектуального інтерфейсу.

Розвиток обчислювальної техніки є базою автоматизації розумової діяльності, і тому виникло нове поняття штучного інтелекту. Проте, успіхи в галузі інтелектуалізації обчислювальної техніки незначні, особливо, якщо порівнювати досягнуте з очікуваними результатами та прогнозами. Доводиться визнавати, що термін «штучний інтелект» відображає поки що тільки нездійсненні очікування. Слід відмітити, що основні проблеми, які зупинили вчених на перших етапах створення мовного інтелектуального інтерфейсу, до цього часу не вирішені.

На даний час проблеми у розвитку систем штучного інтелекту ставлять питання про застосування орієнтації на досягнення високорівневої технології обробки інформації (тобто отримання нової якості), що виявляється у намаганнях реалізувати на фонеймановських комп'ютерах системи ШтІ. Отже сучасні, нові вимоги до технології обробки інформації можна звести до таких двох моментів: необхідність вирішення проблем, що погано формалізуються, та наявність кінцевого користувача, який за своєю професійною діяльністю не є програмістом.

Таким чином приходимо до розуміння одного з варіантів на шляху створення систем ШтІ – це шлях аналізу, моделювання та синтезу мовного інтелектуального інтерфейсу за допомогою алгебри скінченних предикатів (АСП), засобів k -значної логічної системи та k -значних структур і кодування.

Крім цього, ще однією проблемною задачею на сучасному етапі є вирішення завдання створення, розвитку й відродження української науково-тех-

ічної термінології, що через історичні обставини є у значній мірі зрусифікованою, а також спотвореною навіть на рівні елементарної фонетики. Мова ж є однією з головних прикмет нації, а культ нації – другим після релігії джерелом сили народу. Розвиток та широке застосування рідної мови є тим важелем, що формує й породжує національну культуру і науку, виробництво та економіку. Нав'язування і використання чужомовної термінології веде до несприйняття суспільством новітніх технологій, занепаду культури та розпаду державності.

Слід відмітити, що одним із найперспективніших напрямків застосування теорії багатозначних структур і кодування в системах штучного інтелекту на практиці є моделювання природної мови. Цей підхід дозволяє поряд з розвитком варіантів програмної реалізації отриманих моделей мови здійснити інший підхід – схемний. Мозок при цьому розглядається як відправна точка побудови k -значних просторових структур мовних систем штучного інтелекту. Основні результати, що відображують такі дослідження, наведені у [1, 2, 4, 5].

1.2. Постановка проблеми та аналіз останніх досліджень

Якщо машинам призначено на інтелект, то й «міркувати» вони будуть за тими ж законами, що й людина. Нейрофізіологічні дослідження природного інтелекту мозку показують наявність у ньому механізмів багатозначного (k -значного) кодування та просторового характеру активності мереж нервових клітин [1, 2].

Таким чином, зроблено акцент на нейрофізіологічні та нейрокібернетичні аспекти мозкових механізмів. Пов'язано це з тим фактом, що природні нейронні структури із нервових клітин (нейронів) по суті є високоефективними розпізнавальними системами. До того ж ці системи збуджують інтерес не тільки у медичних науковців та фізіологів, а й у фахівців, що займаються питаннями проектування систем штучного інтелекту. Але пряме перенесення результатів дослідження нейрофізіологів до практичної діяльності інженера у даний час неможливе через відсутність відповідної бази. Тому проблеми у розвитку засобів штучного інтелекту, з одного боку, ставлять питання про застосування багатозначного кодування та відповідних структур з просторовими принципами побудови, а з іншого – існують значні проблеми у створенні багатозначних елементів та структур. Відправною точкою у

цих дослідженнях є аналогія поміж багатозначними структурами із просторовими принципами побудови та мікроструктурою кори мозку. Тобто можна провести паралель (описати аналогію) поміж біологічними та технічними схемами.

Мозок — це моделюючий пристрій і більш нічого; це машина широко паралельної дії; це повільнохід, бо кожне нервово волокно може передавати тільки до 1000 імпульсів у секунду. Швидкість руху імпульсів по нервовому волокну незмірно менше, ніж у машини. Тобто усього десятки метрів за секунду, а по дратові машини імпульс рухається зі швидкістю світла — 300 тис. км/с. Дріт здатен передати 10^9 імпульсів за секунду (у сучасних ЕОМ), а може і значно більше (світлові ЕОМ). Мозок має понад 100 млрд. нервових клітин, а кожна з них має приблизно до 10000 нервових волокон [2, 3]. Щільність монтажу елементів в ЕОМ також значна, ступінь мініатюрності — теж. Немає проблеми у створенні ЕОМ з такою ж величезною кількістю елементів. Перевага мозку у його продуктивності. Незважаючи на те, що мозок — повільнохід у порівнянні з ЕОМ, він може дати фори будь-якій ЕОМ. Підраховано, що треба з'єднати 1000 найпотужніших обчислювальних сучасних ЕОМ, щоб одержати інформаційну систему, сумарну за продуктивністю мозку людини. У чому ж перевага мозку? Перевага полягає у просторовому (паралельному) характері його діяльності. А ЕОМ — це система послідовної дії. Отже окремі елементи ЕОМ здатні працювати паралельно й незалежно, як і нейрони мозку. Але цьому заважає принцип програмного управління Джона фон Неймана. Відомо, що ЕОМ здатні працювати за різними принципами: наприклад, машина Т'юрінга працює інакше, але ідея програмного управління у неї присутня.

Як альтернатива, пропонується створення нейроноподібних моделей на базі техніки багатопроекторних обчислювальних систем із програмованою архітектурою, зокрема, на основі цифрових інтегруючих структур [1–5]. Тобто будуються багатозначні інтелектуальні структури (БІС), зберігаючи структуру фон Неймана, які є істотно двозначними. Але за їх допомогою моделюються нейронні процеси просторового підсумовування різних за рівнем впливів нейронних мембран, що мають інерційні та порогові властивості, а також зміни частоти проходження повідомлень, що передаються. Але очевидно, що усі перелічені властивості та функції дискретні за часом та багатозначні за рівнем.

Неадекватність принципів кодування та елементної бази щодо моделюючих процесів, які застосовуються, призводить до надлишковості та ускладнення. Ускладнення, що виникають при створенні БІС, допомагають висунути концепцію адекватності багатозначної логіки, зокрема, у вигляді

АКП, та структур створення БІС за бажаними властивостями та можливостями.

Відомо, що водночас в ЕОМ працює невелика кількість елементів. Паралельно діючі блоки є, але їх дуже мало. Це, наприклад, багаторозрядний підсумувач зі скрізним перенесенням. Але уся логіка працює послідовно, а мозок працює паралельно, тобто усіма нейронами та волокнами. У цьому є його велика перевага. Якби ЕОМ працювала за принципами мозку, то її продуктивність суттєво б збільшилась.

Як стверджується в [1], мозок є однією з відомих нам складних систем переробки інформації. Достатньо сказати, що у ньому знаходиться близько 100 мільярдів нейронів, кожен з яких має, у середньому, 10 000 зв'язків. При цьому мозок надзвичайно надійний: щоденно гине велика кількість нейронів, але він продовжує функціонувати. Обробка великої кількості інформації здійснюється дуже швидко, за частки секунд. Відомо, що нейрон повільнодіючий елемент. Час реагування — кілька мілісекунд.

Розглянемо перспективи побудови ЕОМ за принципами роботи мозку. Так, у ЕОМ водночас імпульси можуть передаватися по 10^2 дратах, а у мозку — по 10^{12} нервовим волокнам (10^{10} клітин і при кожній — по сотні волокон). Маємо вигреш у 10^{10} разів — на розпаралелюванні. У даний час є багатопроекторні ЕОМ (до тисячі процесорів у кращих машинах). Однак все одно залишається вигреш у 10^7 разів за рахунок використання принципів роботи мозку.

Підрахуємо інакше: мозок еквівалентний за потужністю та продуктивністю тисячам найкращих ЕОМ. Одне волокно пропускає до тисячі імпульсів (10^3), а один дріт — 10^9 імпульсів за секунду. Вигреш за рахунок машинізації мозку у 10^6 разів. Так що «машинізований» мозок рівнозначний за потужністю 10^9 ЕОМ. Порядок виграшу приблизно такий.

Водночас у мозку обсяг інформації, яка передається, дорівнює 10^{12} (волокон) $\times 10^3$ (імпульсів за секунду) $= 10^{15}$ (імпульсів за секунду). Але в кращих ЕОМ — 10^3 (каналів) $\times 10^9$ (імпульсів за секунду) $= 10^{12}$ (імпульсів за секунду). Виходить, що треба об'єднати тисячу найкращих ЕОМ, щоб отримати ЕОМ, рівнозначну мозку за продуктивністю.

На сьогоднішній день маємо багатопроекторні ЕОМ, але процесорів у таких ЕОМ не дуже багато (сотні, максимум — тисячі). Рекордна багатопроекторна ЕОМ з рекордною швидкістю роботи була б рівнозначною за продуктивністю мозку. Коли б найкраща багатопроекторна ЕОМ працювала не послідовно, а паралельно, то її продуктивність дорівнювала б 10^{10} мозкам людини, тобто, дорівнювала б об'єднаній інтелектуальній потужності усього людства.

2. Природна мова як об'єкт математичного опису (мовний інтерфейс – частка інтелектуального)

2.1. Аналіз останніх досліджень

Природна мова як явище інтелектуальної діяльності людини є дуже складним об'єктом. Але маючи її формальний опис, його можна реалізувати на ЕОМ і таким чином прищепити машині здібність володіти природною мовою. Треба щоб машина пристосовувалась до людини та розмовляла її мовою, а не навпаки – людина прилаштовувалась до машини. Природна мова для людини найзручніша. Людина ніколи не проміняє її ні на мову математики, ні на мову програмування. Уявляється, що немає іншої галузі знань, яка б у більшій мірі, ніж ця, сприяла підвищенню темпів комп'ютеризації та інформатизації суспільства. Тоді можна буде прищепити машині здібність до природної мови. Якщо можна буде добути з аналізу мови більш досконалу алгебраїчну мову, це дуже збільшить можливості розроблювачів інформаційних систем, нових інформаційних технологій. Алгебро-логічний апарат, який можна буде добувати з природної мови, дозволить розширювати можливості розроблювача, що займається створенням нових інформаційних технологій.

Відомо, що мова з'явилася як особливе відображення паралельної дійсності. Але комп'ютери виникли (історично) у зворотному напрямку: спочатку створювалися і засвоювалися послідовні ЕОМ, а потім почався перехід до паралельних багатопроцесорних обчислювальних систем. Внаслідок цього користувачі-програмісти спочатку оселилися у світі послідовних мов програмування, а зараз перейшли до паралельних мов програмування. Наступний крок на шляху спілкування людини-користувача та ЕОМ – це той ступінь, при якому ЕОМ стане справжнім інтелектуальним помічником, коли вона зможе «зрозуміти» природну мову, тобто момент створення інтелектуального інтерфейсу на підставі природно-мовного інтерфейсу.

Можливо запропонувати наступну узагальнену структуру інтелектуальної системи (ІС), яка задовольняє вимогам нової інформаційної технології рішення задач на ЕОМ. Остання складається з наступних компонентів. Перший компонент – це сукупність засобів інтелектуального інтерфейсу, які мають гнучку структуру, що забезпечує можливість адаптації у широкому спектрі інтересів кінцевих користувачів. Другий – являється сукупністю засобів виконавчої системи, тобто тих програмних та апаратних засобів, які забезпечують обчислювальні, логічні, пошукові та інші аналогічні функції. Третій компонент – це база знань, за допомогою якої здійснюється взаємодія перших двох компонентів (на підставі системи знань, що стосуються проблемного середовища).

Таким чином, основна функція, яку забезпечує інтелектуальний інтерфейс – це спілкування кінцевого користувача з ІС. При цьому використовуються природні для людини форми подання інформації: тексти, зображення, усна мова, а також інші їх комбінації. Однією з головних функцій інтелектуального інтерфейсу є функція підсистеми спілкування на природній мові. Процедури обробки інформації на природній мові утворюють комплекс, що дозволяє як аналізувати, так і синтезувати текст і складають *мовний інтерфейс*.

Відомо, що мова – це система знаків, а мовний знак – це матеріальний носій поняття про предмет. Інакше кажучи, під знаком розуміють матеріальний вираз концепту. Також мова розглядається як засіб програмування розумової діяльності людини. Для ІС зрозуміти значення тексту, який також є видом мовного знаку, означає удавати з себе у формальному вигляді зв'язок між знаком та концептом. З цього випливає наступне: для машинного розуміння тексту необхідно вміти формально уявляти поняття, яке виражене текстом, а також відомості (знання) щодо ситуації сповіщення.

Складність формалізації переходу від знака до його значення полягає у тому, що знаків (текстів) нескінченно багато. Перелічити усі зв'язки між текстами та їх значеннями неможливо. Вихід є – це виявити найпростіші знаки та їх концепти, а потім формалізувати способи отримання концептів складних знаків із концептів простих знаків.

Мова і мислення людини нерозривно пов'язані між собою. Це два взаємодоповнюючих процеси відображення речей (дійсності) у свідомості людини. Мова з'явилася як особливе відображення паралельної реальності і у даний час змогла розвиватися настільки, що сама проявляє тенденцію до незалежного руху та керує діяльністю розуму.

Пізнаючи механізми мови, шляхи її вірного вивчення, людина таким чином намагається розшифрувати та пізнати глибини свого розуму. Такий підхід привів до нового розуміння усієї проблеми. На мову можна дивитися як на деяку алгебраїчну систему. Тобто має право на життя така гіпотеза: природна мова – це деяка алгебра.

2.2. Постановка завдання та його рішення

Отже математичний аналіз механізмів мови дозволяє розглянути її як детермінований, дискретний та скінченний об'єкт. При цьому основна задача полягає у формалізації поняття тексту. Саме поняття тексту теж не просте. Означимо його на рівні інтуїції. Текст розуміємо як математичний опис відношення, яке дозволяє виділити підмножину усіх текстів із деякої множини усіх стрічок, довжина яких обмежена. Таке текстове відношення маємо у вигляді послідовного лінійного ланцюжка символів із букв скінченної абетки. Далі текст

абезпечує
ня кінце-
стовують-
ня інфор-
акож інші
ї інтелек-
и спілку-
обробки
комплекс,
вати текст

а мовний
про пред-
ть матері-
лядається
ності лю-
який та-
ати з себе
знаком та
я машин-
ормально
, а також
ення.

знака до
в (текстів)
ки поміж
хід є — це
и, а потім
онцептів
ків.

пов'язані
х процеси
сті люди-
кнення па-
розвива-
цію до не-
у.

ного вив-
розшиф-
у. Такий
проблеми.
гебраїчну
гіпотеза:

ення

мови доз-
дискрет-
сновна за-
ксту. Саме
о його на
матичний
и підмно-
х стрічок,
відношен-
ланцюж-
Далі текст

сприймаємо як конструкцію з багатьма рівнями. Так, з букв утворюються морфи, з морфів — слова, із слів — словосполучення, з останніх — речення, з речень — абзаци тощо. І далі окремі частини цієї конструкції розглядатимемо як фрагменти тексту. При цьому кожен фрагмент сполучений з оточуючим його текстом та зазнає на собі його впливу. Дію відкинutoї частини тексту замінюють уведенням деякого набору ознак. Останні інтерпретуємо як смисл фрагменту тексту. Відомо, що текст завжди повинен мати цілком визначену думку. Тоді думку, що містить текст, будемо називати смислом.

Таким чином, тексти описують властивості ситуації. Текст відповідає ситуації у тому випадку, якщо він виражає деяку властивість ситуації. При цьому будь-яка думка (це теж властивість ситуації, яких значно більше, ніж ситуацій) формально задається предикатом. Кожна людина являється носієм предикатів тих слів, словосполучень та речень, котрі вона розуміє. Таким чином, смисл речення є предикат, який задає визначений зв'язок $L(X, Y)$ поміж смислом та відповідним йому фрагментом тексту. Тут $X = (x_1, x_2, \dots, x_m)$ змінний вектор смислу, m — число його компонентів, а y_1, y_2, \dots, y_n — деякий фрагмент тексту, де y_1, y_2, \dots, y_n — букви, які стоять на $1, 2, \dots, n$ місцях фрагменту. На підставі наданої моделі базується процес формалізації (математичний опис) природної мови, який запропоновано у роботі.

У якості об'єкту дослідження та моделювання в даній роботі вибрана українська мова (точніше — її словоформи). Для дослідника слово — зовсім не такий простий об'єкт. Воно має складну синтаксичну структуру: з букв утворюються морфи (корені, префікси, суфікси, закінчення), а з них — слова. Сусідні морфи впливають один на одного, що веде до їх взаємної видозміни. Існують жорсткі обмеження на сполучення морфів. Розташування та вибір букв у кожному морфі регулюється досить таки складними законами у залежності від змісту слова. Встановлення послідовності слідування морфів у слові також визначається змістом слова за допомогою складних відношень. Слово зв'язане з оточуючим його текстом багатьма синтаксичними відношеннями, зокрема відношеннями до контексту й до всієї системи знань людини. Окрім літерної (абстрактної) форми слова існує ще й звукове (фізичне) його зображення. Звукова та літерна фор-

ми також зв'язані між собою достатньо складним чином. Усі ці зв'язки в українській мові ще не вивчені інженерною лінгвістикою до кінця.

3. Висновок

Таким чином, пропонується новий принцип побудови надшвидкодійних структур мовних систем штучного інтелекту. В основу функціонування таких систем та структур покладено принцип роботи мозку, що при класичній елементній базі (ВІС, НВІС) найбільш пасує до машин Гільберта [2], ніж до існуючих машин Дж.фон Неймана, Е. Поста, А. Т'юринга, які мають принцип програмного управління. Останній реалізується достатньо повільно. Ось де колосальне підвищення продуктивності: до 10^9 – 10^{10} операцій за секунду. У статті запропонована узагальнена структура інтелектуальної системи, яка відповідає новій інформаційній технології рішення задач на ЕОМ, що орієнтовані на досягнення високорівневої технології обробки мовної інформації (отримання нової якості). Отже, нові вимоги до сучасної технології обробки інформації обумовлені: необхідністю рішення проблем, які погано формалізуються, та наявності користувача, який не є професійним програмістом. Така структура складається з трьох основних компонентів. Мовний інтерфейс розглядаємо як частку інтелектуального на підставі того, що однією з головних функцій інтелектуального інтерфейсу є функція спілкування на природній мові. Перспективи подальших розвідок у даному напрямку полягають у проведенні всебічного аналізу алгебро-логічної структури мови.

Список літератури: 1. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії синтезу надшвидкодійних структур мовних систем штучного інтелекту. К.: ІЗМН, 1997. 264 с. 2. Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Основи теорії багатозначних структур та кодування в системах штучного інтелекту. Харків: Фактор-друк, 2003. 331 с. 3. Бондаренко М.Ф., Осыка А.Ф. Автоматическая обработка информации на естественном языке. К.: УМК ВО, 1991. 142 с. 4. Пат. 2147789 РФ, МКВ Н03К 19/02, Н03М 1/00. Функциональный преобразователь с могозначным кодированием / Бондаренко М.Ф., Коноплянко З.Д., Четвериков Г.Г. Опубл. 24.04.2000. Бюл.11. 4 с. іл. 5. Четвериков Г.Г. Формалізація принципів побудови універсальних k -значних структур мовних систем штучного інтелекту // Доповіді НАН України. 2001, №1 (41). С.76–79.

Надійшла до редколегії 15.11.2004