

РОСТ СКОРОСТИ ПРОГРАММИРОВАНИЯ БИОМЕТРИЧЕСКИХ ПРИЛОЖЕНИЙ ПРИ ИСПОЛЬЗОВАНИИ СПЕЦИАЛЬНЫХ ЯЗЫКОВ АВТОМАТИЧЕСКОГО ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ БОЛЬШОЙ РАЗМЕРНОСТИ

Иванов А.И. (г. Пенза), Язов Ю.К. (г. Воронеж)

В настоящее время системы искусственного интеллекта существенно уступают по своим возможностям человеку. Казалось бы, что наличие более миллиарда транзисторов в современных процессорах и их работа с тактовой частотой 500 ГГц должны давать процессорам огромные преимущества по быстрдействию. Мозг человека содержит примерно столько же нейронов (10 миллиардов нейронов), работающих с тактовой частотой всего 100 Гц (в 5 миллиардов раз медленнее тактовой частоты современных процессоров).

Причиной, по которой люди оказываются умнее современных процессоров, является то, что они способны мгновенно решать своим аналоговым подсознанием задачи огромной размерности [1]. Наш интеллект имеет структуру отображенную на рисунке 1 и состоит на 97% из быстрых аналоговых вычислителей (реализованных на естественных нейронных сетях) и относительно медленного цифрового вычислителя (логического сознания), так же реализованного на естественных нейронах.



Рис. 1 Ускоряюще-уточняющая петля из прямого и обратного многомерных фазовых переходов двух состояний информации

Предположительно, что возможности современного искусственного интеллекта существенно возрастут, если структура его организации будет повторять структуру организации интеллекта человека. При этом речь не идет о реанимации аналоговых вычислителей прошлого века, мы вполне можем заменить их цифровыми эмуляторами нейронных сетей, реализованных программно на обычных процессорах. Получаемый при этом выигрыш обусловлен многократным

переходом информации из неточной аналоговой формы быстрых многомерных вычислений в точную цифровую форму и обратно [1]. Достижимый сегодня выигрыш по уточнению результатов составляет до 12 десятичных порядков, а выигрыш по быстродействию составляет до 40 порядков десятичного представления числа [1] на одном цикле (А/Ц - Ц/А).

Еще одним важнейшим моментом является эффект ускорения программирования биометрических средств искусственного интеллекта за счет перехода к автоматическому обучению двухслойных искусственных нейронных сетей по алгоритмам ГОСТ Р 52633.5 [2]. Традиционное ручное программирование позволяет писать порядка 100 строк программы в день одним человеком. В этом случае на написание ядра Linux (10^7 строк) потребуется 10^5 человеко-дней или 278 человеко-лет. Собрав команду из 278 программистов можно написать проект типа Linux за один год и примерно еще один год уйдет на его тестирование и исправление выявленных ошибок. При этом ручное написание программ может быть выполнено только после полной формализации проекта в виде блок-схем отдельных программ. Как показывает практика для очень многих задач искусственного интеллекта до сих пор не удавалось получить их полную формализацию.

Ситуация коренным образом изменяется, если воспользоваться автоматическим обучением искусственных нейронных сетей алгоритмами ГОСТ Р 52633.5 [2]. В этом случае полная формализация задачи не нужна. Достаточно задать 20 примеров образа «Свой» и порядка 100 примеров случайных образов «Чужие». Далее осуществляется синтез случайных связей между нейронами двухслойной сети. Так эмулятор нейросети изделия «Персона» создает двухслойную сеть с 256 нейронами в каждом слое. Каждый нейрон имеет по 24 входа, то есть программный эмулятор имеет $2 \cdot 256 \cdot 24 = 12\,288$ связей (строк программы). Программирование этих связей через обучение при реализации на обычной ПЭВМ требует порядка 0.33 секунды. То есть мы имеем скорость автоматизированного программирования порядка 3 700 строк в секунду, что быстрее ручного программирования в 10^7 (десять миллионов) раз. Это означает, что подобный автоматизированный способ программирования позволит создавать большие проекты типа Linux силами одного программиста в течении одного рабочего дня. Мы выходим на фантастические скорости программирования, осуществляемого без полной формализации задачи.

Еще одним важным требованием к новой технологии является необходимость иметь очень высокий уровень доверия к ней. Каждое из совокупности нейросетевых решений после его программирования должно быть проверено тестами по ГОСТ Р 52633.3 [3]. Время на каждую из таких проверок уходит примерно столько же, как и на обучение. При этом полностью цифровая (логическая) часть проекта (3% от объема) подобному тестированию не поддается и ее проверку придется все таки осуществлять традиционными способом. В этом отношении помочь может создание автоматических комментаторов логической (полностью формализованной) части программ искусственного интеллекта [5]. То есть, скорость автоматического синтеза и тестирования псевдо аналоговой части программ может быть очень высокой, а вот скорость написания цифровой части программ и скорость ее тестирования пока остаются прежними.

К сожалению, все перечисленные выше преимущества достижимы пока только при создании биометрических приложений защиты информации. Эти приложения имеют хорошо отработанную цифровую часть в виде реализации того или иного протокола криптографической аутентификации. Если же пытаться перенести этот подход на другие приложения искусственного интеллекта, то возникает острая необходимость в создании специального языка осуществляющего взаимодействие высокоразмерного аналогового подсознания и цифрового

искусственного сознания гораздо меньшей размерности. Новый язык программирования двух совершенно разных фрагментов искусственного интеллекта с одной стороны должен наследовать свойства традиционных формальных языков программирования цифровой части, однако с другой стороны он должен уметь работать с многомерным псевдо аналоговым нейросетевым подсознанием. То есть все новые языки должны иметь подобно естественным языкам [5] некоторую итерационную структуру постоянного уточнения результатов подсознательного аналогового мышления и итерационную структуру ускоренного поиска (сопоставления между собой) уже известных нейросетевых решений.

На данный момент такой итерационный язык еще не создан. Существующие стандартизованные интерфейсы [6] взаимодействия между аналоговой и цифровой частью приложений и протоколы взаимного объединения нескольких биометрических технологий [7] следует рассматривать, как первые робкие попытки создания чрезвычайно обедненного (упрощенного) варианта языка для создания и эксплуатации только биометрических приложений. Тем не менее эти первые попытки дают фантастический рост скорости программирования в сравнении с традиционным ручным написанием программ. Это связано прежде всего с тем, что цифровая часть биометрических приложений (криптографические протоколы аутентификации) стандартизована и не нуждается в переписывании и тестировании. Цифровая часть нейросетевого искусственного интеллекта биометрических приложений защиты информации одинакова для огромного многообразия вариантов реализации псевдо аналоговой нейросетевой предобработки.

В иных приложениях искусственного интеллекта может оказаться, что цифровую часть придется переписывать для каждого нового класса задач. В этом случае получить миллиардные ускорения процессов программирования нельзя. В этом случае ускорение получается примерно в 33 раза, так как 3% от общего объема программ необходимо по прежнему переписывать в ручную.

Литература:

1. Иванов А.И. Подсознание искусственного интеллекта: программирование автоматов нейросетевой биометрии языком их обучения. Пенза-2012, электронная книга издательства ОАО «ПНИЭИ», 125 с.
2. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».
3. ГОСТ Р 52633.3-2011 «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора».
4. Пустыгин А.П., Иванов А.И., Язов Ю.К., Соловьев С.В. Автоматический синтез комментариев к программным кодам: перспективы развития и применения «Программная инженерия» № 3 2012 г., с. 30-34.
5. Хомский Н. Язык и мышление. М.: Издательство МГУ, 1972 г.
6. ГОСТ Р 52633.4 «Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия-код».
7. ГОСТ Р 52633.7 «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация».

Материалы поступили 25.05.2012, опубликовано в Интернет 05.06.2012 по положительной рецензии д.т.н., профессора Малыгина А.Ю. (Пенза).