



Извлечение правил из классических, глубоких нейронных сетей и нейро-нечетких сетей¹

А. Н. Аверкин (averkin2003@inbox.ru)

Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия, Учебно-научная лаборатория искусственного интеллекта, нейротехнологий и бизнес-аналитики, РЭУ им. Г. В. Плеханова, Москва, Россия

Аннотация

В этой статье делается попытка дать обзор нескольких алгоритмов извлечения правил из искусственной нейронной сети. Второй целью этой статьи является установление фундаментальных связей между двумя важными областями искусственного интеллекта — нечеткой логикой и глубоким обучением. Такой подход позволит исследователям в области нечеткой логики разрабатывать прикладные системы в области сильного искусственного интеллекта, которые также представляют интерес для специалистов в области машинного обучения.

Ключевые слова: Глубокое обучение, Нейронные сети, Извлечение правил, Конволюционные нейронные сети, Машинное обучение, Искусственный интеллект

ВВЕДЕНИЕ

В этой статье рассматриваются основные понятия машинного обучения, которые имеют отношение к контексту извлечения правил из классических и глубоких нейронных сетей. Она включает проблему классификации методов обучения на основе правил из нейронных сетей. Затем мы рассмотрим текущее состояние извлечения правил из нейронных сетей. Последняя часть обсуждает конкретные проблемы при работе с глубокими нейронными сетями и нейро-нечеткими сетями. На данном этапе мы также предлагаем некоторые алгоритмы, которые могут успешно извлечь правила из этих более сложных нейронных сетей

Искусственные нейронные сети (ANN) являются широко известными параллельными вычислительными моделями, которые демонстрируют отличное поведение в решении сложных проблем искусственного

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 20-07-00770 А).

интеллекта. Тем не менее, многие исследователи отказываются использовать их из-за того, что они являются «черным ящиком». Это означает, что определение того, почему нейронная сеть принимает именно такое решение, является трудной задачей.

Это существенный недостаток, так как в этом случае трудно доверять надежности сети, которая решает ответственные практические проблемы. Например, это имеет место в критических, с точки зрения безопасности, приложениях, где скрытая ошибка может привести к опасным для жизни действиям или огромным экономическим потерям.

Кроме того, изучение того, как нейронные сети извлекают, хранят и преобразуют знания, может быть полезно для будущих методов машинного обучения. Например, повышение прозрачности нейронных сетей может помочь обнаружить так называемые скрытые зависимости, которые не присутствуют во входных данных, но появляются в результате их интеграции в нейронную сеть. Чтобы преодолеть этот недостаток нейронных сетей, исследователям пришла в голову идея извлечения правил из нейронных сетей, которые могут стать связующим звеном между символическими и коннекционистскими моделями представления знаний в искусственном интеллекте.

Большинство авторов сосредотачиваются на извлечении наиболее понятных правил, и в то же время они должны имитировать поведение нейронной сети как можно точнее, вплоть до изоморфного представления нечетких правил в виде нейро-нечеткой системы. С 1992 года, с момента публикации докторской диссертации Джанга по нейро-нечетким сетям, в этой области была проделана большая работа, которая закончилась созданием направления мягких вычислений. С тех пор было разработано и проанализировано большое количество методов извлечения правил из нейронных сетей, и были получены интересные результаты в различных подходах. Однако несмотря на то, что существует довольно много известных алгоритмов, ни один из них никогда не был явно протестирован в глубоких нейронных сетях. Кроме того, большинство авторов сосредотачиваются на сетях, в которых имеется лишь небольшое количество скрытых слоев.

Только в последние несколько лет появилась новаторская работа по анализу конкретных методов извлечения правил из глубоких сетей и представлены новые алгоритмы, способные выполнить эту задачу.

1. Методы извлечения правил из нейронных сетей

В искусственном интеллекте нейронные сети и методы обучения на основе правил являются двумя подходами к решению проблем классификации. Оба метода являются известными вариантами моделей обучения, которые предсказывают классы для новых данных. Для многих задач нейросетевые методы обучения, основанные на правилах, являются очень точными.

Тем не менее, нейронные сети имеют один главный недостаток: способность понимать понятийную сущность обученных моделей у нейросети слабее, чем у подходов, основанных на правилах. Концепции, полученные при обучении нейронных сетей, трудно понять, потому что они представлены с использованием большого набора параметров [Craven, 1994].

Повышение прозрачности нейронных сетей путем извлечения из них правил имеет два основных преимущества. Это дает пользователю некоторое представление о том, как нейронная сеть использует входные переменные, чтобы принять решение позволяет выявить скрытые функции в нейросетях, когда правила используются для объяснения отдельных нейронов. Выявление особо важных атрибутов или выявление причин ошибок нейронной сети может быть частью понимания. Пытаясь сделать непрозрачные нейронные сети более понятными, методы извлечения правил устраняют разрыв между точностью и ясностью [Johansson, 2006], [Craven, 1999].

Для того чтобы, например, нейронная сеть использовалась в критически важных приложениях, таких как самолеты и электростанции, требуется более понятная форма. В этих случаях крайне важно, чтобы у пользователя системы была возможность проверить выход искусственной нейронной сети при всех возможных условиях входа [Andrews, 1995].

Для формализации задачи извлечения правил из нейронной сети, можно использовать следующее определение : «При заданных обученной нейронной сети и данных, на которых она была обучена, создайте описание сетевой гипотезы, которая понятна, но приближается к поведению заданной сети».

Чтобы различать различные подходы к извлечению правил из нейронных сетей, в [Andrews, 1995] была введена многомерная таксономия. Первое измерение, которое в ней описывается, является выразительной силой извлекаемых правил (например, правила IF-THEN правила или нечеткие продукционные правила).

Второе измерение называется прозрачностью и описывает стратегию, за которой следует алгоритм извлечения правил. Если метод использует нейронную сеть только как черный ящик, независимо от архитектуры нейросети, мы называем его педагогическим подходом. Если вместо этого алгоритм учитывает внутреннюю структуру нейронной сети, мы называем этот подход декомпозиционным. Если алгоритм использует компоненты как педагогических, так и декомпозиционных методов, то этот подход называется эклектическим.

Третьим измерением является качество извлеченных правил. Поскольку качество является широким термином, оно делится на несколько критериев, а именно: аккуратность, точность, непротиворечивость и понятность. В то время как аккуратность измеряет способность правильно классифицировать ранее не видимые примеры, точность измеряет степень, в которой правила хорошо могут имитировать поведение нейронной сети [Johansson, 2006].

Точность может рассматриваться как точность по отношению к выходу нейронной сети. Непротиворечивость может быть измерена только тогда, когда алгоритм извлечения правил включает обучение нейронной сети вместо обработки уже обученных нейронной сети. Извлеченный набор правил считается непротиворечивым, когда нейронная сеть

генерирует наборы правил, которые правильно классифицируют тестовые данные для различных сессий обучения. Понятность рассматривается здесь как мера размера правил, то есть, короткие правила при их небольшом количестве считаются более понятным.

В этой статье мы сосредоточимся только на трех описанных критериях. В соответствии с [Thrun, 1993], мы ориентируемся на методы, которые не предъявляют особых требований к тому, как была обучена нейронная сеть до того, как были извлечены правила. Кроме того, анализируются только алгоритмы, способные извлекать правила из нейронных сетей прямого распространения, несмотря на любые другие характеристики архитектуры. В соответствии [Craven, 1999] мы хотим, чтобы алгоритм обладал наибольшей степенью общности.

Давайте проанализируем некоторые методы извлечения правил, которые отвечают вышеуказанным характеристикам. Начнем с декомпозиционного подхода. Как упоминалось ранее, декомпозиционные подходы для извлечения правил из нейронных сетей действуют на уровне нейронов. Как правило, декомпозиционный подход анализирует каждый нейрон, и формирует правила, которые имитируют поведение этого нейрона. Мы рассмотрим ниже КТ алгоритм, полиномиальный алгоритм Цукимото и экстрактор правил через индукцию дерева решений.

Алгоритм КТ был одним из первых подходов к разложению для извлечения правил из нейронных сетей [Fu, 1994]. Алгоритм КТ описывает каждый нейрон (слой за слоем) с правилами IF-THEN путем эвристически поиска комбинаций входных атрибутов, превышающего порог нейрона. Модуль перезаписи используется для получения правил, относясь к исходным атрибутам ввода, а не к выводам предыдущего уровня. Чтобы найти подходящие комбинации, метод КТ применяет поиск на дереве, то есть правило (представленное в качестве узла в дереве) на этом уровне генерирует свои детские узлы, добавляя дополнительный доступный атрибут. Кроме того, алгоритм использует ряд эвристик, чтобы остановить рост дерева в ситуациях, когда дальнейшее улучшение невозможно

Полиномиальный алгоритм Цукимото для извлечению правил из нейронной сети очень похож на метод КТ. Он также использует многоуровневый декомпозиционный алгоритм для извлечения правил IF-THEN для каждого нейрона, а также отслеживает стратегию поиска входных конфигураций, превышающих порог нейрона. Основным преимуществом алгоритма Цукимото является его вычислительная сложность, которая является полиномиальной, в то время как метод КТ является экспоненциальным. Алгоритм достигает полиномиальной сложности, ища соответствующие термины, используя пространство многолинейных функций. На втором этапе эти термины используются для создания правил IF-THEN. Впоследствии, если таковые имеются, обучающие данные используются для повышения точности правил. На последнем этапе алгоритм Цукимото пытается оптимизировать понятность удаляя из правил несущественные атрибуты.

Другой метод извлечения правил путем индукции дерева принятия решений был введен в [Sato, 2001]. Их алгоритм CRED преобразует каждую выходную единицу нейронной сети в решение, где узлы деревьев тестируются с помощью узлов скрытого слоя, а листья представляют класс. После этого из этого шага извлекаются промежуточные правила. Затем для каждой точки разветвления, используемой в этих правилах, другое дерево решений создается с помощью точки разветвления

на входном слое нейронной сети. В новых деревьях листья не выбирают непосредственно класс. Извлечение правил из второго дерева принятия решения приводит нас к описанию состояния скрытых нейронов, состоящих из входных переменных. В качестве заключительного шага заменяются промежуточные правила, описывающие выходной слой через скрытый слой на те, которые описывают скрытый слой на основе входов нейронной сети. Затем они объединяются, чтобы построить правила, описывающие выход нейронной сети на основе ее входных данных.

Педагогические подходы не учитывают внутреннюю структуру нейронной сети. Целью педагогических подходов является рассмотрение обученных нейросетей, как целостного объекта или как черного ящика [Tickle, 1998]. Главная идея состоит в том, чтобы извлечь правила, непосредственно сопоставляя входные данные с выходными данными [Setiono, 2000].

Педагогические подходы имеют доступ имеют дело с нейронной сетью, как с функцией. Эта функция задает выход нейронной сети для произвольного входа, но не дает понимания внутренней структуры нейронной сети или ее весов. Для заданной нейронной сети этот класс алгоритмов пытается найти связь между возможными вариациями ввода и выходами, созданную нейронной сетью, причем некоторые из них используют определенные обучающие данные, а некоторые нет.

Извлечение правил, основанное интервальном анализе, использует анализ доверительных интервалов (VIA) для извлечения правил, имитирующих поведение нейронных сетей. Основная идея этого метода заключается в том, чтобы найти входные интервалы, в которых выходной сигнал нейронной сети стабилен, то есть прогнозируемый класс одинаков для слегка меняющихся конфигураций ввода. В результате анализ доверительных интервалов обеспечивает основу для надежных корректных правил.

Получение правил с использованием выборки представляет собой несколько методов, которые следуют более или менее той же стратегии для извлечения правил из нейронной сети с помощью выборки, то есть, они создают обширный набор данных в качестве основы для извлечения правил. После этого выбранный набор данных представляется в стандартный алгоритм обучения для генерации правил, имитирующих поведение сети. В [Johansson, 2006] доказано, что использование выборочных данных эффективнее, чем использование только обучающих данных в проблемах извлечения правил.

Одним из первых методов, последовавших за этой стратегией, был алгоритм Трепана [Tickle, 1998]. Он работает аналогично алгоритму «разделяй и властвуй» выполняя поиск точек разветвления на обучающих данных для отдельных экземпляров разных классов. Основными отличиями от «разделяй и властвуй» являются лучшая стратегия расширения древовидной структуры, дополнительные точки разветвления и возможность выбора дополнительных примеров обучения в более глубоких узлах дерева. В результате алгоритм также создает дерево решений, которое, однако, может быть преобразовано в набор правил, если это необходимо.

Алгоритм, извлечения правил двоичного входа и выхода (BIO-RE), способен обрабатывать только NN с двоичными или бинаризованными входными атрибутами. BIO-RE создает все возможные комбинации входных данных и запрашивает их у нейронной сети. С помощью вывода NN для каждого примера создается таблица истинности. От таблицы истинности также легко перейти к правилам, если это необходимо.

ANN-DT является еще одним методом выборки, основанном на принятии решений на для описания поведения нейронной сети. Общий алгоритм основан на алгоритме CART с некоторыми вариациями в первоначальной реализации. ANN-DT использует метод выборки для расширения, с тем чтобы большая часть обучающей выборки была репрезентативной. Это достигается с помощью метода ближайшего соседа, в котором рассчитывается расстояние от точки выборки до ближайшей точки в наборе обучающих данных» обучающей выборки и сравнивается с эталонным значением.

Идея создания большого набора примеров на первом этапе также реализуется алгоритмом STARE. Как и BIO-RE, STARE также формирует обширные таблицы истинности для обучения. Преимущество STARE заключается в его способности не только обрабатывать двоичные и дискретные атрибуты, но и работать с непрерывными входными данными. Для формирования таблиц истинности алгоритм перестраивает входные данные, в то время как для каждого непрерывного атрибута необходимо выбирать его во всем д Примером педагогического подхода с использованием выборки образовательных данных, которые мы хотим представить, является KDRuleEx. Как и Трепан, этот алгоритм также генерирует дополнительные обучающие примеры, в случае, когда данных для следующих точек разделения слишком мало. KDRuleEx использует генетический алгоритм для создания новых примеров обучения. Эта техника приводит к таблице принятия решений, которая может быть преобразована, например, в правила IF-THEN.

Эклектический подход — это методы извлечения правил, включающие элементы как педагогического, так и декомпозиционного подхода. В частности, эклектический подход использует знания о внутренней архитектуре и вектора весов векторов в нейронной сети в дополнение к символьному алгоритму обучения использует знания о внутренней архитектуре и вектора весов векторов в нейронной сети в дополнение к символьному алгоритму обучения.

Подход быстрого поиска правил в нейронной сети включает в себя подход FERNN, который сначала пытается определить соответствующие скрытые нейроны, а также соответствующие входы в сеть. Для этого шага строится дерево решений с использованием алгоритма С4.5. Процесс извлечения правил приводит к генерации правил M-of-N и IF-THEN. Имея набор правильно классифицированных примеров обучения, FERNN анализирует значения активации каждой скрытой вершины. Для каждой скрытой вершины значения активации сортируются в порядке возрастания. Затем используется алгоритм С4.5, чтобы найти наилучшую точку разделения для формирования дерева решений.

2. Извлечение правил из глубоких нейронных сетей и нейро-нечетких сетей

Наиболее интересным с точки зрения данного исследования является извлечение правил с использованием нейро-нечетких моделей. Системы, основанные на нечетких

правилах (FRBS), разработанные с использованием нечеткой логики, стали полем активных исследований за последние несколько лет. Эти алгоритмы доказали свои сильные стороны в таких задачах, как управление сложными системами, создание нечетких элементов управления. Взаимоотношения между обоими мирами (ANN и FRBS) были тщательно изучены и показана их эквивалентность [Setiono, 2000]. Это дает два немедленных и важных вывода. Во-первых, мы можем применить то, что было обнаружено для одной из моделей, к другой. Во-вторых, мы можем перевести знания, встроенные в нейронную сеть, на более когнитивно-приемлемый язык — нечеткие правила. Другими словами, мы получаем семантическую интерпретацию нейронных сетей [Averkin, 2018a], [Pilato, 2018], [Averkin, 2018b].

С 2012 года началась революция сетей глубокого обучения. Рассмотрим одну из первых и, пожалуй, самых цитируемых работ — глубокую сверточную сеть Alexnet. Она был первым чемпионом по распознаванию образов. Она имеет 7 скрытых слоев, 650 000 нейронов, 60 000 000 параметров. Училась на двух GPU в течение 1 недели. Где мы получим достаточно изображений, чтобы обучить ее?

В 2010 году появился набор данных Imagenet из 15000000 образов. Появление Imagenet подняло изучение нейронных сетей на совершенно новый уровень. Параллельно быстро развивались вычислительные мощности, которые привели компьютерное зрение к тому виду, который мы знаем и любим сейчас. С 2010 года также проводится ежегодный конкурс Imagenet, где впервые в 2012 году победила сверточная нейронная сеть Alexnet, и с тех пор глубокие сети не утратили своих позиций. Последний победитель, National Assembly, представленная учеными из Китая, содержала 269 слоев.

Для того, чтобы получить семантическую интерпретацию черного ящика глубокого обучения нейро-нечетких сети могут быть использованы вместо последнего полносвязного слоя. Например, ANFIS (адаптивная нейро-нечеткая система) является многослойной сетью прямого распространения. Эта архитектура имеет пять слоев, таких как нечеткий слой, продукционный слой, слой нормализации, слой дефаззификации и выходной слой. ANFIS сочетает преимущества нейросети и нечеткой логики.

Цель комбинации нечеткой логики и нейронных сетей состоит в том, чтобы создать архитектуру, которая использует нечеткую логику, чтобы явно показать знания фантастическим образом, в то время как обучающая природа нейронной сети максимизирует ее параметры. ANFIS, разработанная Джангом в 1992 году, интегрирует преимущества как нейронной сети, так и нечетких систем, которые не только обладают хорошими возможностями обучения, но и могут быть легко интегрированы. ANFIS используется во многих областях, таких как аппроксимация функций, интеллектуальное управление и прогнозирование временных рядов [Boyacioglu, 2010].

Сети глубокого обучения и нейро-нечеткие сети могут быть объединены различными способами. Гипотетическая система может быть создана

с использованием двух компонентов. Первый — это генерация функций глубокого обучения, которые могут быть использованы для создания репрезентативных функций из текста напрямую. Система глубокого обучения первоначально будет обучена немаркированным данным. Как только эти функции будут извлечены из системы глубокого обучения, они будут интегрированы в систему нечеткого вывода. Эта система может включать как функции, обнаруженные в результате глубокого обучения, так и субъективную информацию от аналитиков. Эти две части вместе могут быть использованы для целей классификации. Таким образом, окончательная система сможет сообщать как о результатах классификации, так и о конкретных особенностях и правилах, которые были активированы для того, чтобы система пришла к своему заключению. Кроме того, окончательная система может быть дополнительно откорректирована аналитиком в качестве одной из форм обратной связи.

Благодаря теории нечетких множеств, используя нечеткие отношения и правила, можно создать эффективную модель прогнозирования временных рядов с большим числом входов и одним выходом (прогнозом). Подобный подход позволяет проиллюстрировать в некотором роде обоснование работы искусственной нейронной сети, используя с одной стороны нейро-нечеткие модели, а с другой нечеткие когнитивные карты. Нами была разработана гибридная модулярная модель прогнозирования, объединяющая теорию нечеткой логики, когнитивных карт и искусственных нейронных сетей.

Модулярная система в целом состоит из нескольких специализированных модулей. В общем случае, эти модули имеют следующие характеристики:

1. Модули системы специфичны и имеют специализированные вычислительные архитектуры, чтобы распознавать и реагировать на определенные подзадачи большой общей задачи.
2. Каждый модуль, как правило, не зависит от других модулей в его функционировании и не влияет на работу других модулей.
3. Модули имеют более простую архитектуру, по сравнению с системой в целом. Таким образом, модуль работает быстрее, чем сложная монолитная система.
4. Результаты работы каждого модуля по отдельности объединяются с помощью специального интеграционного модуля (в нашем случае модуль консенсус прогноза), благодаря чему достигается наиболее высокая точность прогноза всей системы.

Система имеет три основных модуля, отвечающих за задачу прогнозирования. Нейро-нечеткая сеть ANFIS, выполняет прогноз временного ряда на основе численных показателей и дает нам, так называемый количественный прогноз, результаты которого проходят через систему верификации (оценки адекватности прогноза), если прогноз соответствует необходимой точности, то он передается на следующий модуль. Параллельно с нейро-нечеткой сетью, работает модуль с нечеткой когнитивной картой, который получает на вход данные о событийном влиянии на временной ряд, строится когнитивная карта, в которой учитываются все факторы влияния на конкретный прогнозируемый показатель. На выходе когнитивная карта дает нам прогноз с вероятностью его выполнения, то есть, с консонансом фактора, который говорит нам, выполнится прогноз или нет. Далее все данные,

полученные с данных модулей поступают на третий модуль, работающий на основе ANFIS сети, который агрегирует информацию, полученную с предыдущих модулей и выдает итоговый консенсус прогноз. На рис. 1 представлена модель системы прогнозирования.



Рис.1. Модульная система прогнозирования

В этой статье делается попытка дать обзор нескольких алгоритмов извлечения правил из искусственной нейронной сети. Некоторые из современных алгоритмов обсуждаются в каждой категории — декомпозиционные, педагогические и эклектические. В настоящее время глубокое обучение обеспечивает приемлемое решение для множества проблем. Это новая область машинного обучения, которая, как считается, продвигает машинное обучение на шаг вперед. Но это все еще система черного ящика. Только несколько работ пытаются установить фундаментальную связь между двумя важными областями в искусственном интеллекте, то есть глубоким обучением и нечеткой логикой.

Изучение нечеткой логики достигло кульминации в конце 20-го века, и с тех пор начало уменьшаться. Это снижение может быть частично связано с отсутствием результатов в машинном обучении. Извлечение правил является одним из способов помочь понять нейронные сети. Эти исследования проложат путь для исследователей нечеткой логики для разработки приложений в области искусственного интеллекта и решения сложных проблем, которые также представляют интерес для сообщества машинного обучения. Опыт и знания в области нечеткой логики хорошо подходят для моделирования неоднозначностей в больших данных, моделирования неопределенности в представлении знаний и обеспечения обучения передаче с неиндуктивным выводом

Список литературы

1. [Craven, 1994] Craven, M. and Shavlik, J. W. (1994). Using sampling and queries to extract rules from trained neural networks. In ICML, pages 37–45.
2. [Johansson, 2006] Johansson, U., Lofstrom, T., Konig, R., Sonstrod, C., and Niklasson, L. (2006). Rule extraction from opaque models—a slightly different perspective. In Machine Learning and Applications, 2006. ICMLA'06. 5th International Conference on, pages 22–27.
3. [Craven, 1999] Craven, M. and Shavlik, J. (1999). Rule extraction: Where do we go from here. University of Wisconsin Machine Learning Research Group Working Paper, pages 99–108.
4. [Andrews, 1995] Andrews, R., Diederich, J., and Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-based systems, 8(6): 373–389.

5. [Thrun, 1993] Thrun, S. (1993). Extracting provably correct rules from artificial neural networks. Technical report, University of Bonn, Institut für Informatik III.
6. [Fu, 1994] Fu, L. (1994). Rule generation from neural networks. *Systems, Man and Cybernetics, IEEE Transactions on*, 24(8): 1114–1124.
7. [Sato, 2001] Sato, M. and Tsukimoto, H. (2001). Rule extraction from neural networks via decision tree induction. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, volume 3, pages 1870–1875.
8. [Tickle, 1998] Tickle, A. B., Andrews, R., Golea, M., and Diederich, J. (1998). The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks*, 9(6): 1057–1068.
9. [Setiono, 2000] Setiono, R. and Leow, W. K. (2000). FERNN: An algorithm for fast extraction of rules from neural networks. *Applied Intelligence*, 12(1–2):15–25.
10. [Averkin, 2018a] Averkin A., Yarushev S. Hybrid Neural Networks and Time Series Forecasting. *Artificial Intelligence. Communication in Computer and Information Sciences 934* — Springer, 2018 — pp. 230–239
11. [Pilato, 2018] Giovanni Pilato, Sergey A. Yarushev, and Alexey N. Averkin Prediction and Detection of User Emotions Based on Neuro-Fuzzy Neural Networks in Social Networks // *Proceedings of the Third International Scientific Conference «Intelligent Information Technologies for Industry» (IITI'18) Volume 2* — Springer. *Advances in Intelligent Systems and Computing*, volume 875. pp. 118–26
12. [Averkin, 2018b] Averkin, G Pilato and S. A. Yarushev. An Approach for Prediction of User Emotions Based on ANFIS in Social Networks A. N. // *Second International Scientific and Practical Conference Fuzzy Technologies in the Industry, FTI 2018– CEUR Workshop Proceedings, 2018*, pp. 126–134
13. [Boyacioglu, 2010] Boyacioglu, Melek Acar, and Derya Avci. «An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange.» *Expert Systems with Applications* 37.12 (2010): 7908–7912.

RULES ACQUISITION FROM CLASSIC, DEEP AND NEURO-FUZZY SYSTEMS

A. N. Averkin (averkin2003@inbox.ru)

Federal Research Centre of Informatics and Computer Science of RAS, Moscow, Vavilova, 42, Moscow, Russia, Educational and Scientific Laboratory of Artificial Intelligence, Neuro-technologies and Business Analytics, Plekhanov Russian University of Economics, Moscow, Russia

S. A. Yarushev (Yarushev.SA@rea.ru)

Plekhanov Russian University of Economics, Moscow, Russia

Abstract

This paper attempts to give an overview of several algorithms for extracting rules from an artificial neural network. The second goal of this article is to establish fundamental connections between two important areas of artificial intelligence — fuzzy logic and deep learning. Such an approach will allow researchers in the field of fuzzy logic to develop applied systems in the field of strong artificial intelligence, which are also of interest to specialists in the field of machine learning.

Keywords: Deep learning, Neural networks, rule extraction, Convolutional neural network, Machine learning, Artificial intelligence.