

Нейро-символический искусственный интеллект в коллаборативных системах поддержки принятия решений*

А. В. Смирнов, А. В. Пономарев, Н. Г. Шилов, Т. В. Левашова

Санкт-Петербургский Федеральный исследовательский центр РАН, Санкт-Петербург, Россия

Аннотация. В статье рассмотрены требования к коллаборативным человеко-машинным системам поддержки принятия решений и основные проблемы, возникающие при их создании. Показана роль методов нейро-символического искусственного интеллекта для обеспечения коммуникации между разнородными участниками коллаборативной системы поддержки принятия решений. Проведен анализ современных результатов в области онтолого-ориентированного нейро-символического искусственного интеллекта, в первую очередь, нацеленных на объяснение нейросетевых моделей с помощью онтологий и использование символических знаний для повышения эффективности нейросетевых моделей. Предложена концептуальная модель коллаборативной человеко-машинной системы поддержки принятия решений на основе онтолого-ориентированного нейро-символического интеллекта.

Ключевые слова: системы поддержки принятия решений, нейро-символический интеллект, коллаборативные системы.

DOI 10.14357/20718594220303

Введение

Развитие технологий искусственного интеллекта (ИИ) и растущая сложность принятия решений при управлении сложными динамическими системами обуславливают необходимость совместной работы человека и ИИ, в том числе, в составе коллективов разнородных участников (например, экспертов и агентов, действующих на основе ИИ).

В последние годы нейросетевой подход к реализации ИИ совершил существенный скачок, связанный с развитием идей так называемого «глубокого обучения». В ряде задач глубокие нейронные сети позволяют получать результаты, незначительно уступающие возможностям человека, а в отдельных случаях и

превосходят его. Тем не менее, практическое применение глубоких нейронных сетей в ряде приложений (и, в частности, в коллаборативных системах поддержки принятия решений) сдерживается недостатком интерпретируемости результатов, неявным характером области входных воздействий, при которых результат оказывается надежным, что закономерно приводит к снижению доверия к ним [1]. Кроме того, применение глубоких нейронных сетей требует использования значительных объемов обучающих данных, что ограничивает их применение в задачах, где их сбор затруднен.

Поскольку взаимодействие людей в значительной мере опирается на использование символов, методы нейро-символического интеллекта представляются наиболее адекватным

* Работа выполнена при финансовой поддержке РФФ, проект 22-11-00214.

✉ Пономарев Андрей Васильевич. E-mail: ponomarev@iias.spb.su

решением для построения эффективных коллаборативных систем поддержки принятия решений (СППР), существенным образом зависящих от обмена информацией, формирования доверия и прочих процессов, характерных для сообществ людей [2]. Одной из перспективных формальных символьных систем является представление знаний о проблемной области с использованием онтологий.

В целом, теория и практика сотрудничества между человеком и современным ИИ (основанным, в том числе, на использовании глубоких нейронных сетей) только начинает формироваться, о чем говорит достаточно большое количество теоретических и методологических статей в этой области, но относительно малое число практических решений.

1. Требования к коллаборативным человеко-машинным системам поддержки принятия решений и основные проблемы их создания

В основе коллаборативных СППР, построенных на базе сотрудничества искусственного интеллекта и человека, лежит теория социотехнических систем. Она описывает, как социальные и технические системы должны совместно работать, отдавая приоритет координации навыков, знаний и талантов человека и ИИ, в отличие от чисто технических систем, в которых вначале разрабатывается технический компонент, впоследствии приспособляемый к потребностям человека [3]. Все виды коллаборативных человеко-машинных СППР являются подклассом социотехнических систем. Коллаборативные СППР, построенные на базе сотрудничества ИИ и человека, предъявляют строгие требования к наличию в них возможностей для обучения человека от ИИ и к адаптации ИИ к потребностям человека [4].

Требования к коллаборативным СППР на основе нейро-символического искусственного интеллекта объединяют в себе требования ко всему классу человеко-машинных СППР и к коллаборативным СППР, построенным на базе сотрудничества ИИ и человека [5–7]. Ниже в перечне требований ИИ обозначает любой цифровой объект (агент, сервис, приложение и т.п.), обладающий искусственным интеллектом.

1. ИИ и человек должны быть взаимно предсказуемы в своих действиях. В частности, ИИ должен быть осведомлен о склонностях человека (например, к предвзятости).

2. ИИ и человек должны в текущей ситуации иметь представление друг о друге, задаче и окружающей среде. Здесь предполагается, что ИИ и человек должны иметь представление о характеристиках друг друга (например, текущая рабочая нагрузка; предпочтения, имеющиеся знания; эмоциональное состояние; компетенции; история принятия предвзятых суждений и т.д.) и о работе в целом.

3. ИИ и человек должны быть взаимно управляемы. Данное требование подразумевает, что ИИ и человек могут брать на себя и делегировать друг другу задачи. Это позволяет ИИ, например, принимать и делегировать задачи и обязанности с учетом возможных ограничений, а также понимать динамические изменения в заданиях, выданных человеком. ИИ получает возможность самостоятельно предлагать или даже объявлять о таких изменениях или привлекать внешние экспертные знания. Кроме того, ИИ может вмешиваться в процесс принятия решений, чтобы предотвратить нежелательные или опасные действия со стороны человека.

4. ИИ и человек должны иметь представление о контексте, в котором принимаются решения, и о своей роли в этом процессе.

5. ИИ и человек должны понимать друг друга. Человек должен понимать, какой вклад может внести ИИ, а какой нет, понимать, почему ИИ получил тот или иной результат; он должен уметь формулировать свои потребности в форме или на языке, понятном ИИ. ИИ должен уметь предвидеть потребности человека, используя знания о модальностях, объяснять свои действия и решения, сообщать об ошибках и стараться пояснять их причины, доказывать, что человек может доверять ИИ, взаимодействовать с человеком на понятном ему языке (в понятной ему модальности).

6. ИИ и человек должны уметь обучаться друг у друга посредством взаимодействия и обратной связи в процессе принятия ими решений.

7. ИИ и человек должны быть предварительно обучены взаимодействию друг с другом.

8. СППР должна предоставить «прозрачный» интерфейс для взаимодействия ИИ-человек,

а также представление информации о текущей ситуации, которое понятно ИИ и человеку.

В настоящее время системы, основанные на ИИ, все чаще используются для решения широкого класса задач в различных сферах человеческой деятельности (например, распознавание и генерация образов, системы персонализации). Поддержка принятия решений не является исключением. Однако сотрудничество подразумевает не только совместную деятельность в какой-либо сфере для достижения общих целей и согласия (консенсуса), но также обучение и обмен знаниями. И здесь на пути создания коллаборативных человеко-машинных систем есть две фундаментальные проблемы:

1. Обучение систем ИИ в процессе взаимодействия с людьми сегодня не является высокоэффективным, в том числе по причине недостаточной проработанности методов генерализации (обобщения) на основе малых данных и очень ограниченных способностей методов машинного обучения по использованию априорных знаний.

2. Системы, основанные на глубоких нейронных сетях, обладают слабой интерпретируемостью генерируемых ими решений, что затрудняет их использование в ряде задач (медицинская диагностика, управление предприятием) – там, где последствия принимаемых решений оказывают серьезное влияние на здоровье или благосостояние человека [8]. Данная проблема обостряется при применении нейронных сетей в составе коллаборативных СППР, основанных на взаимодействии людей-

экспертов и ИИ-агентов, действующих на основе ИИ. Подотчетность, прозрачность, интерпретируемость играют особую роль в коллаборативных системах [9, 10] и исторически большинство коллаборативных СППР (и групповых СППР) уделяли этому аспекту особенное внимание [11].

Таким образом, для решения этих важнейших проблем необходимо решить следующие задачи (Рис. 1):

1. Разработать способы обеспечения объяснимости моделей ИИ, использующих знания о проблемной области, выраженные в форме онтологий. Терминология проблемной области, закодированная в машиночитаемом виде, допускающим определенные рассуждения (касающиеся принадлежности индивидов классам) является перспективным инструментом объяснения по своей природе – это одновременно и терминология, используемая людьми, и строгая модель, основанная на логике.

Существующие исследования дают определенные свидетельства и в пользу того, что результат работы нейронной сети может быть снабжен онтологическим объяснением [12], и в пользу того, что подобное объяснение может быть более понятно пользователю, чем другие виды объяснений [13]. Тем не менее, исследования в данном направлении находятся еще на начальном этапе и на большую часть вопросов только предстоит получить ответы. В частности, актуальными являются следующие вопросы:

- Могут ли существующие онтологии быть использованы для объяснения уже существующих нейросетевых моделей (т.н. постериорное

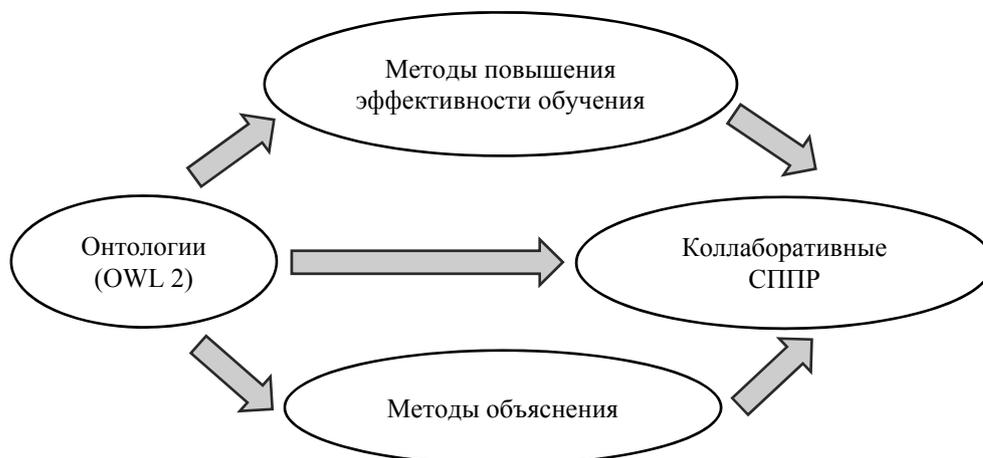


Рис. 1. Взаимосвязь задач

объяснение)? Какие условия для этого должны выполняться (возможные сочетания цели ИНС, структуры онтологии, характера объяснения)? Как это сделать эффективно?

- Как онтологии могут быть использованы при формировании изначально объяснимых нейросетевых моделей, обеспечивая при этом минимальную деградацию качества предсказаний (или без подобной деградации)?

- Как измерять степень полезности объяснений?

2. Разработать методы использования априорных знаний в форме онтологии для повышения эффективности нейросетевых моделей машинного обучения (в частности, снижение требовательности к объему данных, используемых для обучения, и/или повышение точности при использовании сопоставимых объемов данных). При этом следует рассмотреть применение выраженных в форме онтологий знаний на всех этапах разработки и использования модели машинного обучения: подготовка исходных данных, формирование архитектуры и гиперпараметров нейронной сети, интерпретация результатов.

3. Создать технологические основы построения коллаборативных СППР, основанных на синтезе символического и нейросетевого подходов к реализации ИИ. Решение этой задачи позволит создать основы для формирования эффективного партнерства между человеком-экспертом или лицом, принимающим решение (ЛПР) и агентом, действующим на основе ИИ, в том числе, в коллаборативных СППР.

Комплексное решение перечисленных задач послужит основой для организации систем, основанных на объяснимом ИИ и способных обучаться в процессе взаимодействия с экспертами-людьми для поддержки принятия решений. Технологические аспекты функционирования коллаборативных человеко-машинных систем в значительной степени опираются на результаты исследований о применимости тех или иных способов объяснения моделей ИИ в различных ситуациях, выводы о структуре онтологий и другие. Решение трех данных задач может быть обобщено в набор рекомендаций и практик, касающийся как создания таких гибридных систем (методы управления знаниями и специфические методы онтологического моделирования, упрощающие конвергенцию онтологического и нейросетевого подходов, методы

создания и обучения нейросетевых моделей), так и их функционирования (методика объяснения, эффективные механизмы взаимодействия между участниками коллаборативной системы принятия решений и другие).

2. Анализ современных результатов в области онтолого-ориентированного нейро-символического ИИ

Существует два основных подхода к организации систем ИИ – символический и нейросетевой. Каждый из них обладает своими достоинствами и недостатками. Так, символический подход характеризуется относительно высокой интерпретируемостью моделей, но и высокими трудозатратами на инженерию знаний. Нейросетевой – достаточно высокой точностью представления различных задач и отсутствием необходимости в инженерии знаний, однако высокими требованиями к количеству обучающих примеров и низкой интерпретируемостью. Их успешная интеграция позволит сгладить недостатки, присущие данным подходам, и расширить спектр применения технологий ИИ, в первую очередь, посредством снижения требований к объему данных и повышения интерпретируемости (объяснимости) создаваемых моделей. Одним из современных и перспективных направлений, где интеграция символического и нейросетевого подходов к реализации ИИ оказывается особенно востребованной, являются коллаборативные СППР, включающие как людей-экспертов, так и программные сущности, основанные на ИИ, и предназначенные для решения определенного класса задач. Применительно к СППР, подобная интеграция существенно улучшит двустороннюю коммуникацию между экспертами и ИИ-агентами (перенос выраженных с помощью символов знаний и опыта экспертов ИИ-агентам и получение объяснений решений ИИ-агентов) в процессе их совместной работы, что может существенно повысить эффективность как ИИ-агентов, так и всей СППР.

Методы нейро-символического интеллекта являются довольно широкой областью исследований, активно развивающейся на протяжении последних 15 лет. В данной статье рассматривается один из видов символического

подхода к ИИ (представление знаний о проблемной области в виде онтологий) и показывается, как представленные таким образом знания могут быть использованы для улучшения и объяснения результатов, получаемых с помощью моделей глубокого обучения (многослойных нейросетей).

Использование знаний, выраженных в форме онтологий, для улучшения моделей глубокого обучения. Идея совместного использования онтологического описания проблемной области с методами машинного обучения появилась достаточно давно, хотя до настоящего времени и не получила большого внимания со стороны научного сообщества. Так, еще в 2015 году была опубликована работа, предложившая объединение технологии сверточных нейронных сетей с онтологиями для вывода абстрактных закономерностей в изображениях индийских памятников [14]. Перенос накопленных знаний (transfer learning), при котором знания о проблемной области передаются в нейронную сеть во время обучения (сверху вниз), а вывод делается с использованием предсказания сети и онтологической таксономии (снизу вверх), позволил существенно повысить качество модели по сравнению с простым нейросетевым классификатором. Авторами было сделано предположение, что объединение онтологии и сверточных нейронных сетей может привести к лучшим результатам на сложных и больших наборах данных, где простые сверточные нейронные сети оказываются недостаточно эффективными.

Развитие данной идеи можно увидеть, например, в работе [15], посвященной классификации звуков. Авторы работы разрабатывают таксономию звуков и дополняют ее соответствующими классификаторами. Результаты демонстрируют эффективность предлагаемого решения как для выявления отношений в онтологии, так и для повышения точности классификации звуков.

Схожий подход предложен в работе [16]. Сначала основанная на ограниченной машине Больцмана (Restricted Boltzmann Machine) нейронная сеть обучается выделять шаблоны поведения на основе их описания в онтологии (атрибутов и отношений с другими паттернами), а затем она используется для классификации поведения на основе заданных атрибутов. Такой подход позволил не только существенно

повысить точность модели определения шаблонов поведения, но и получить объяснимую модель.

Несколько иной подход предложен в работе [17], где используются много достаточно простых нейросетевых классификаторов, иерархически организованных на основе онтологии, что приводит к повышению точности классификации (в том числе и за счет отдельно обученных моделей, ориентированных на анализ объектов определенных, например, трудноразличимых, классов).

Дальнейшее развитие подобных подходов происходило в области «ручного» описания признаков классифицируемых объектов, выявляемых более простыми моделями машинного обучения, и создания онтологических моделей, позволяющих классифицировать объекты на основе данных признаков. В то время как экстенциональное («классическое») обучение заключается в представлении алгоритму как можно большего количества экземпляров концепции, интенциональная информация концентрируется на основных характеристиках, которые кратко отражают характеристики концепции. Такие подходы позволили существенно уменьшить требуемые для обучения объемы данных, сделать модели машинного обучения более устойчивыми, а также повысить степень их объяснимости. Так, в работе [18], ориентированной на решение задачи классификации рукописных цифр, в качестве признаков были использованы графические примитивы (наличие линий и кружков на изображении). Комбинация экстенционального и интенционального обучения позволила авторам достичь такой же точности, как и экстенциональное обучение, но с использованием только одной шестой исходного набора обучающих данных (на наборе MNIST). В работе [19] онтологическая модель описывает связи и свойства элементов сообществ микробов в зависимости от среды их обитания. Авторами была получена высокая точность классификации при использовании существенно меньшей обучающей выборки, чем обычно требуется «классическим» моделям машинного обучения.

Использование символьных знаний для обеспечения объяснимости и интерпретируемости моделей глубокого обучения. Существует несколько постановок задачи объяснения предсказаний, получаемых с помощью

моделей ИИ. Первый критерий их классификации связан с тем, объясняется ли вся модель (например, аппроксимация логики модели с помощью хорошо интерпретируемых деревьев решений [13]) или результат предсказания для одного конкретного образца (например, указанием на то, какие именно составные части – признаки – образца тем или иным образом повлияли на решение). Другим важным в практическом смысле критерием является подразделение на так называемые *post-hoc* (апостериорные) методы и методы построения *self-explainable* (самообъясняемых) моделей. Апостериорные методы нацелены на объяснение уже существующих моделей, обученных без предъявления специфических требований к объяснимости, в то время как методы построения самообъясняемых моделей представляют определенные требования к структуре моделей машинного обучения (в частности, глубокого обучения), чтобы обеспечить их объяснимость.

Особенностью апостериорных методов является то, что основная модель-предсказатель обучается без учета возможности объяснения. То есть данные методы объяснения применимы (потенциально) к любой, уже обученной ИНС.

В основе большинства апостериорных методов онтолого-ориентированного объяснения предсказаний нейронных сетей лежит аппроксимация нейронной сети более интерпретируемой моделью (например, деревом решений) – либо в окрестности заданного примера (для локальных методов объяснения), либо в целом (для глобальных методов). Онтологии (и свойства тех или иных концептов) при этом могут использоваться как в ходе формирования самого дерева [13] (использование концептов при генерации разбиений, помещение более общих концептов и их признаков выше в дереве решений), так и при управлении процессом обучения модели-аппроксиматора [20] (выбор образцов, семантически схожих с тем, для которого формируется объяснение).

В относительно недавней статье [12] предлагается альтернативный способ апостериорного объяснения предсказаний нейронных сетей, заключающийся в определении соответствия между возбуждением нейронов сети и концептами онтологии. Авторы показали, что формирующиеся в ходе обучения (например, сверточной) нейронной сети внутренние представления часто можно сопоставить концептам онтологии, свя-

занным с целевой меткой (например, через которые целевая метка может быть определена). Подобные соответствия между активациями основной сети и концептами онтологии могут быть установлены с помощью специальных нейронных сетей-«декодеров», каждая из которых соответствует одному концепту и обучается как бинарный классификатор, на вход которого поступают активации основной сети для определенного примера, а выходом является факт принадлежности данного примера определенному концепту. Таким образом можно получить множество концептов, связанных с примером, и сформировать объяснение как вывод с помощью онтологии на основе распознанных концептов.

Онтолого-ориентированные самообъяснимые модели, как правило, предполагают, что каждому из концептов онтологии, которые будут участвовать в формировании объяснений, назначается отдельный нейрон [21–24]. В большинстве случаев этот нейрон является одним из выходных нейронов сети, то есть множество выходных данных нейронной сети дополняется множеством концептов онтологии, актуальных для примера, и нейронная сеть предсказывает не только класс примера, но и предоставляет набор логических значений, которые могут быть использованы машиной логического вывода при построении объяснений. В ряде случаев предсказание целевой переменной вообще формируется машиной вывода (полнофункциональной или упрощенной, представленной как набор операций для вычисления на графическом процессоре [24]) либо линейной моделью над выходами нейронной сети, обеспечивая строгое соответствие между самим предсказанием и его объяснением. Концепты могут соответствовать и внутренним нейронам сети. Так, в [21] концепты онтологии упорядочиваются по уровню общности и распределяются по слоям полносвязной нейронной сети в зависимости от этого уровня; при обучении полученной сети с помощью специальной функции потерь накладывается штраф на связи, не соответствующие онтологии.

В целом, экспериментальные исследования [13, 20] показывают, что использование знаний проблемной области, закодированных в виде онтологии, повышает качество объяснений.

Основные выводы по анализу исследований:

1. Объяснимость является важным свойством моделей ИИ (в особенности, при построении коллаборативных систем человек-ИИ

[25]), при этом многие исследователи проводят разграничение между интерпретируемостью и объяснимостью. Интерпретируемость при этом трактуется просто как математическая связь между входами и выходами системы-«черного ящика» (например, LIME), а объяснимость – человеческое понимание логики и/или семантики такой системы [26].

2. Научные публикации, в которых объяснение предсказаний нейросетевых моделей ИИ достигается с помощью применения онтологий, начинают появляться, однако они носят частный характер и ориентированы, как правило, на решение локальных задач. Это делает особенно востребованными и актуальными фундаментальные исследования в данном направлении.

3. Интеграция семантических признаков (например, выраженных с помощью онтологий) для обеспечения объяснения и повышения точности моделей машинного обучения посредством использования априорной информации являются связанными задачами, однако отсутствуют публикации, которые бы рассматривали их комплексно (и совместно). Вместе с тем, некоторые исследования показывают, что интеграция семантических признаков в нейросетевые модели и лучшая объяснимость ведут к улучшению качества моделей машинного обучения [26].

3. Перспективные направления онтолого-ориентированного нейро-символического интеллекта

Общая схема онтолого-ориентированного нейро-символического интеллекта приведена на Рис. 2. Ее основу составляют три процесса –

онтологическое моделирование, нейросетевое моделирование и представление результатов. В ходе онтологического моделирования происходит выделение наиболее значимых концептов предметной области и фиксация их описаний и свойств посредством языка представления онтологий [27]. Основным результатом этого процесса является собственно онтология, выраженная на том или ином языке, например, OWL 2. Процесс нейросетевого моделирования нацелен на подбор весов нейросетевой модели таким образом, чтобы обеспечить наилучшее значение функции ошибок на заданной выборке. Ключевыми этапами здесь являются подготовка исходных данных для обучения, выбор архитектуры нейронной сети и способа обучения, а также обработка результатов. Особенностью онтолого-ориентированного нейро-символического интеллекта является то, что онтология, являющаяся результатом онтологического моделирования, тем или иным образом используется на каком-либо из перечисленных этапов (или на нескольких). Данное отношение использования выделено на схеме пунктирными стрелками. Разработка конкретных методов использования знаний, закодированных в онтологии, на том или ином этапе нейросетевого моделирования для повышения качества модели или для обеспечения ее объяснимости задает структуру исследований в области онтолого-ориентированного нейро-символического интеллекта.

В данном разделе перечислены основные направления исследований, раскрывающие взаимосвязи, обозначенные на приведенной схеме (например, способы учета знаний, представленных в виде онтологии, на тот или иной этап нейросетевого моделирования). Эти направле-

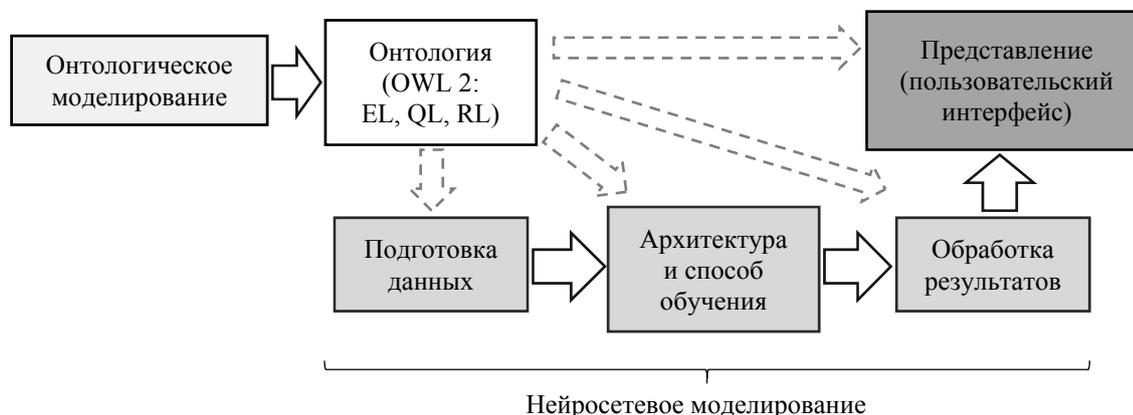


Рис. 2. Схема онтолого-ориентированного нейро-символического интеллекта

ния сгруппированы по двум аспектам или двум важнейшим целям исследований: обеспечение объяснимости результатов, полученных с помощью нейросетевых моделей, и повышение качества нейросетевых моделей за счет использования знаний, выраженных в виде онтологий.

Онтолого-ориентированное объяснение нейросетевых моделей. Для реализации объяснения целесообразно рассмотреть различные постановки задачи объяснения нейросетевых моделей (в частности, апостериорное объяснение «черного ящика» и формирование изначально объяснимых моделей) и расширить их использованием знаний о проблемной области, выраженных с помощью онтологий. Так, при апостериорном объяснении необходимо рассмотреть различные методы определения соответствия между активациями нейронов объясняемой сети и связи обрабатываемого образца с теми или иными концептами онтологии. В современных работах, где применяется подобный подход [12], обучается одна объясняющая сеть на каждый концепт, что оказывается достаточно неэффективно. Одним из возможных улучшений является использование предварительно рассчитанных вложений (embeddings) концептов онтологии, сохраняющих семантическую близость, и решение задачи отображения между векторами вложений и активациями промежуточного слоя объясняемой сети. При этом необходимо рассмотреть различные разновидности ИНС – сверточные, рекуррентные, сети прямого распространения, смешанные.

Необходимо провести систематизацию ситуаций, в которых целесообразно объяснение, с учетом целей, требований к объяснению, существенных характеристик самой основной задачи (решение которой должно быть объяснено). Систематизация может опираться, например, на работу [28], но учитывать специфику семантических (онтолого-ориентированных) объяснений. Затем для выделенных ситуаций следует выбрать наиболее перспективные методы формирования объяснений. Подобная систематизация, в совокупности с рекомендациями, может быть использована в качестве научно-обоснованного руководства человеком, но также может быть преобразована в машиночитаемый вид и использована в составе автоматизированного средства, помогающего принимать решения об использовании тех или иных мето-

дов нейро-символического интеллекта при разработке СППР.

Важным компонентом разработок в области объяснимого ИИ является сам интерфейс пояснения. То есть, разрабатывается не только определенный класс моделей, которые допускают объяснение, но и их способ их представления в рамках пользовательского интерфейса. Этот интерфейс (работающий, как правило, в тесной связи с объясняемой моделью) позволяет, ответить пользователю на все вопросы: почему так, каковы ограничения (когда модель будет работать, а когда нет), можно ли доверять решению.

Наличие устоявшихся и общепринятых способов оценки и сравнения эффективности подходов является одним из условий быстрого прогресса в области исследований (в частности, сосредоточенность на определенном наборе «бенчмарков» является одним из серьезных факторов, обусловивших развитие методов глубокого обучения в последнее десятилетие). Таким образом, необходимо разработать способ оценки семантических (онтолого-ориентированных) объяснений нейросетевых моделей. Сложность данной задачи обусловлена тем, что сама цель объяснения тесно связана с человеком (для которого объяснение формируется), соответственно, и достоверная оценка объяснения может быть получена только с привлечением человека (или нескольких, для снижения субъективности оценки). Однако необходимость привлечения людей существенно замедляет цикл разработки и валидации методов объяснения. Соответственно, целесообразно разработать комплекс методик – для оффлайн и онлайн-оценки методов объяснения, обладающих существенной корреляцией. Наличие оффлайн-метода (основанного на теоретико-информационных представлениях и анализе структуры онтологии) позволит упростить прототипирование методов семантического объяснения и сократить трудозатраты на оценку методов объяснения с привлечением людей.

Использование априорных знаний для повышения эффективности нейросетевых моделей ИИ. Одним из перспективных подходов является использование онтологии для формирования архитектуры ИНС, например, с помощью внедрения в нее специальным образом предобученных слоев, кодирующих концепты онтологии, их отличительные признаки и связи между ними.

Вместе с тем, необходимо рассмотреть не только методы использования априорных знаний, предполагающие их учет в самой архитектуре нейросетевой модели или методе обучения (например, конструирование специальной функции потерь, учитывающей онтологические отношения), но и проанализировать возможности, относящиеся к различным этапам цикла построения и применения модели – сбору и подготовке данных (включая разметку), определению структуры модели и метода ее обучения, а также обработке результатов работы модели (так называемой постобработке).

На основе экспериментов с различными онтологиями необходимо исследовать значимость тех или иных онтологических структур (выражений) на объяснимость моделей и ценность соответствующих видов априорной информации для повышения эффективности нейросетевых моделей. Здесь целесообразно сфокусировать внимание на языке онтологий OWL 2 и его профилях (OWL EL, OWL QL, OWL RL).

4. Концептуальная модель

На основе результатов проведенного анализа литературы по организации коллаборативных систем (в том числе, основанных на человеко-машинном сотрудничестве), построена обобщенная концептуальная модель коллаборативной человеко-машинной СППР (Рис. 3).

Исходной точкой для функционирования такой СППР является взаимодействие между ли-

цом, принимающим решение (ЛПР), и командой, состоящей из людей-экспертов и ИИ-агентов. ЛПР формулирует проблемную ситуацию и для поиска возможных ее решений (и их оценки) формируется команда, включающая как людей (экспертов), так и ИИ-агентов (программные компоненты, как правило, реализующие какие-то интеллектуальные функции обработки информации и способные принимать участие в коллективной работе над задачей). ЛПР может принимать участие в формировании и координации работы команды. Команда, в свою очередь, формируется динамически под проблемную ситуацию. При этом, существенными характеристиками команды являются [29]: модель организации команды (схема подчинения), способы распределения задач и ответственности.

Работа над проблемной ситуацией декомпозируется в решение частных задач и интеграцию их результатов. Основные типы таких задач выделены, например, в работе [30]: распознавание, предсказание, решение и действие. Данные типы могут быть использованы, в частности, в описании ИИ-агентов, для облегчения автоматизированного распределения задач.

Конкретная команда для поддержки принятия решения при работе с заданной проблемной ситуацией формируется в рамках выбранной модели организации команды из экспертов и ИИ-агентов, каждый из которых характеризуется набором возможностей по решению тех или иных задач. Характеристики команды (мо-

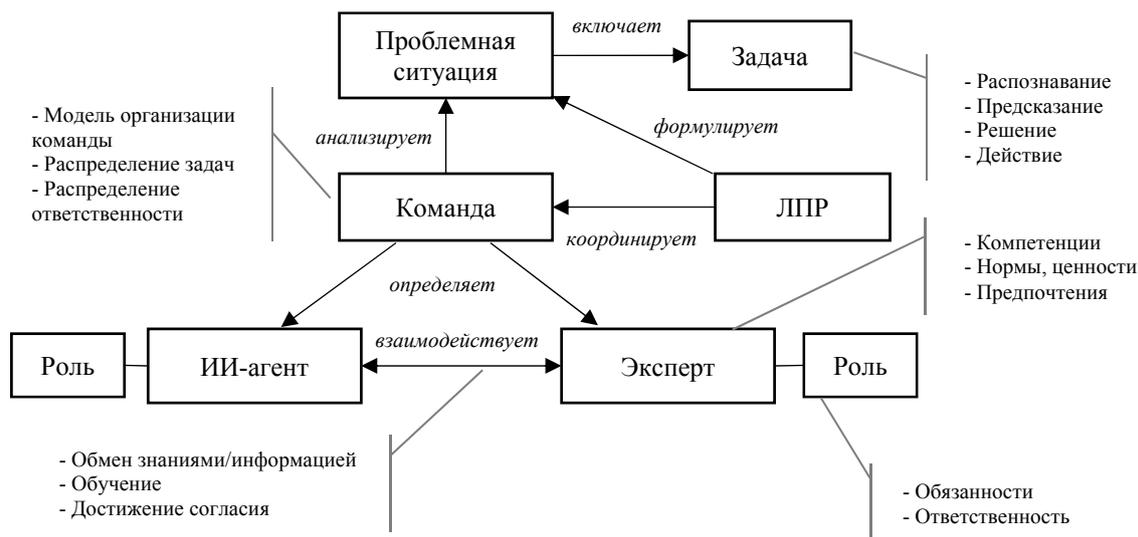


Рис. 3. Обобщенная концептуальная модель коллаборативной СППР

дель организации, распределения задач и ответственности), определяют ее состав и правила взаимодействия участников. Они используются как при формировании (например, подборе участников, определении ролей), так и транслируются участникам в виде норм и правил взаимодействия, которые могут носить как неформальный характер, так и подкрепляться функциями среды, обеспечивающей совместную работу. Участие в команде, как правило характеризуется ролью, определяемой как совокупность обязанностей по решению тех или иных задач и ответственности за результат.

Ключевыми базовыми процессами, обеспечивающими работу над проблемой, являются обмен информацией и знаниями (в результате которого снижается общая неопределенность проблемной ситуации), обучение, понимаемое как передача норм, знаний, и навыков, а также достижение согласия между участниками команды. Конкретные механизмы реализации этих процессов определяют итоговый дизайн коллаборативной СППР.

На основе описанной концептуальной модели предложена схема коллаборативной СППР, основанной на онтолого-ориентированном ней-

ро-символическом интеллекте (Рис. 4). В предлагаемой схеме конкретизируется часть общих процессов, выделенных в обобщенной концептуальной схеме коллаборативным СППР (обмен информацией, обучение и другие).

В общем случае ИИ-агенты могут реализовывать любую логику и быть основаны на различных технологиях ИИ, однако в данной статье рассматриваются, в первую очередь, ИИ-агенты, основанные на нейросетевом подходе.

В основе предлагаемой схемы коллаборативной СППР лежит представление проблемной ситуации (и прочей информации, имеющей отношение к принятию решения) с помощью онтологии. При этом целесообразным оказывается использование нескольких онтологий, раскрывающих различные аспекты проблемы: терминология и закономерности предметной области, структура задачи принятия решения как таковой, логика аргументации и пр. Данные аспекты могут проблемно-ориентированным образом соединяться в единую онтологию с помощью аппарата многоаспектных онтологий [31, 32].

Представленная таким образом информация о проблеме содержится в онтолого-ориентированном интеллектуальном пространстве [33, 34]



Рис. 4. Схема коллаборативной СППР, основанной на онтолого-ориентированном нейро-символическом интеллекте

(доступная участникам системы «классная доска», хранящая множество высказываний в терминах онтологии и обеспечивающая удобные механизмы доступа и оповещения), к которому имеют доступ как ИИ-агенты, так и эксперты. Достоинством онтолого-ориентированного представления является то, что информация оказывается доступна как людям (напрямую или через специализированный интерфейс), так и ИИ-агентам, которые могут использовать возможности языка SPARQL для извлечения релевантных высказываний из описания текущего состояния проблемы.

Взаимодействие между участниками коллектива, осуществляющего поддержку принятия решений, происходит опосредованно – через изменение содержимого интеллектуального пространства.

В контексте проблематики данной статьи, центральную роль в общей схеме СППР играют ИИ-агенты. С точки зрения осуществляемых преобразований информации можно выделить две категории таких агентов. Агенты первой группы осуществляют преобразования символической информации на основе знаний, то есть, и на входе, и на выходе у этих агентов символическая информация (выраженная как набор высказываний на онтологии предметной области). Агенты же второй группы осуществляют преобразование «сырых» данных в символическую информацию. На вход таким агентам попадают изображения, видео-, аудиоданные или просто численные наблюдения, на выходе же у них высказывания, основанные на онтологии предметной области. Для организации подобных агентов в настоящее время применяются различные технологии и, в частности, ИНС, с помощью которых подобные «сырые» данные могут быть отнесены к одному из концептов. Именно в агентах этой группы оказываются востребованными технологии нейро-символического интеллекта. Такие агенты используют нейросетевую парадигму для интерпретации данных в символической (понятийной) системе, что оказывается необходимым для последующей коммуникации.

Одним из существенных аспектов коллаборативной поддержки принятия решений является отслеживаемость (provenance), то есть возможность отследить источник для каждого утверждения (касающегося перечня альтернатив или их оценок), которые сформулированы в ходе совместной работы над задачей. Подобная

возможность играет ключевую роль, например, в формировании и использовании механизмов доверия, существенно снижающих накладные расходы на верификацию. В предлагаемой концептуальной модели отслеживаемость обеспечивается двумя механизмами. Во-первых, все изменения в представлении проблемы снабжаются метаданными в соответствии с онтологией PROV-O [35], позволяющей описать источники информации (участника системы, предоставившего информацию, основания). Во-вторых, агенты помимо самого результата работы добавляют и объяснение – почему именно такой результат был получен. Для агентов, осуществляющих преобразование символической информации, – это цепочка логического вывода, а для агентов, осуществляющих преобразование «сырых» данных в символические – это те или иные виды онтолого-ориентированного объяснения.

Заключение

В статье рассмотрены проблемы создания коллаборативных СППР, основанных на сотрудничестве человека и ИИ. Одним из перспективных способов обеспечения коммуникации между разнородными участниками таких систем являются методы нейро-символического искусственного интеллекта. В частности, онтологические представления, строгим образом фиксирующие терминологию проблемной области и позволяющие сформировать объяснение в понятных эксперту терминах.

Проведен анализ современных результатов в области онтолого-ориентированного нейро-символического искусственного интеллекта, в первую очередь, нацеленных на объяснение нейросетевых моделей с помощью онтологий и использование символических знаний для повышения эффективности нейросетевых моделей. Выделены перспективные направления исследований. Предложена концептуальная модель коллаборативной человеко-машинной системы поддержки принятия решений на основе онтолого-ориентированного нейро-символического интеллекта.

Литература

1. Аверкин А.Н. Объяснимый искусственный интеллект: итоги и перспективы // Интегрированные модели и мягкие вычисления в искусственном интеллекте

- (ИММВ-2021). Сборник научных трудов X-й Международной научно-технической конференции. 2021. С. 153–174.
2. Демидовский А.В., Бабкин Э.А. Интегрированные нейросимволические системы поддержки принятия решений: проблемы и перспективы // Бизнес-информатика. 2021. Т. 15, № 3. С. 7–23.
 3. Appelbaum S.H. Socio-technical systems theory: an intervention strategy for organizational development // *Manag. Decis.* 1997. Vol. 35, No. 6. P. 452–463.
 4. Rafner J. et al. Deskillung, Upskilling, and Reskilling: a Case for Hybrid Intelligence // *MORALS + MACHINES*. 2021. P. 24–39.
 5. Johnson M. et al. Coactive Design: Designing Support for Interdependence in Joint Activity // *J. Human-Robot Interact.* 2014. Vol. 3, No. 1. P. 43.
 6. Bosch K. van den, Bronkhorst A. Human-AI Cooperation to Benefit Military Decision Making: STO-MP-IST-160. Soesterberg. 2018. 13 p.
 7. Xiong W. et al. Challenges of human—machine collaboration in risky decision-making // *Front. Eng. Manag.* 2022. Vol. 9. No. 1. P. 89–103.
 8. Уткин Л.В. и др. Обзор методов объяснения и интерпретации принятия решений в интеллектуальных системах диагностики онкологических заболеваний // *Искусственный интеллект и принятие решений*. 2020. № 4. P. 55–65.
 9. Cummings M.L. Automation and Accountability in Decision Support System Interface Design // *J. Technol. Stud.* 2006. Vol. 32. No. 1.
 10. Kroon M.B.R., Van Kreveld D., Rabbie J.M. Group Versus Individual Decision Making: Effects of Accountability and Gender on Groupthink // *Small Gr. Res.* 1992. Vol. 23. No. 4. P. 427–458.
 11. *Mastering Data-Intensive Collaboration and Decision Making* / ed. Karacapilidis N. Cham: Springer International Publishing. 2014. Vol. 5.
 12. Ribeiro de Sousa M., Leite J. Ali gning Artificial Neural Networks and Ontologies towards Explainable AI // *Proceedings of the AAI Conference on Artificial Intelligence*. 2021. Vol. 35, No. 6. P. 4932–4940.
 13. Confalonieri R. et al. Using ontologies to enhance human understandability of global post-hoc explanations of black-box models // *Artif. Intell.* Elsevier B.V. 2021. Vol. 296. P. 103471.
 14. Gupta U., Chaudhury S. Deep transfer learning with ontology for image classification // 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). IEEE. 2015. P. 1–4.
 15. Sun Y., Ghaffarzagdegan S. An Ontology-Aware Framework for Audio Event Classification // *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020. P. 321–325.
 16. Phan N. et al. Ontology-based deep learning for human behavior prediction with explanations in health social networks // *Inf. Sci. (Ny)*. 2017. Vol. 384. P. 298–313.
 17. Kuang Z. et al. Integrating multi-level deep learning and concept ontology for large-scale visual recognition // *Pattern Recognit.* 2018. Vol. 78. P. 198–214.
 18. Fong A.C.M., Hong G. Ontology-Powered Hybrid Extensional-Intensional Learning // *Proceedings of the 2019 International Conference on Information Technology and Computer Communications - ITCC 2019*. New York, New York, USA: ACM Press. 2019. P. 18–23.
 19. Zha Y. et al. Ontology-Aware Deep Learning Enables Ultrafast, Accurate and Interpretable Source Tracking among Sub-Million Microbial Community Samples from Hundreds of Niches // *bioRxiv*. 2020.
 20. Panigutti C., Perotti A., Pedreschi D. Doctor XAI: An ontology-based approach to black-box sequential data classification explanations // *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020. P. 629–639.
 21. Bourgeais V. et al. Deep GONet: self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data // *BMC Bioinformatics*. BioMed Central. 2021. Vol. 22. P. 1–24.
 22. Voogd J. et al. Using Relational Concept Networks for Explainable Decision Support // 3rd IFIP Cross Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE 2019. LNC 11713. 2019. P. 78–93.
 23. Abbass H.A. et al. Machine Education: Designing semantically ordered and ontologically guided modular neural networks // 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE. 2019. P. 948–955.
 24. Bourguin G. et al. Towards Ontologically Explainable Classifiers // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2021. Vol. LNCS 12892. P. 472–484.
 25. Neerinx M.A. et al. Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance // 15th International Conference on Engineering Psychology and Cognitive Ergonomics, EPCE 2018 Held as Part of HCI International 2018, LNAI 10906. Springer. 2018. P. 204–214.
 26. Donadello I., Dragoni M. SeXAI: A Semantic Explainable Artificial Intelligence Framework // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2021. Vol. 12414, June. P. 51–66.
 27. Загорулько Ю.А. Современные средства формализации семантики областей знаний на основе онтологий // *Информационные и математические технологии в науке и управлении*. 2018. Vol. 11, No. 3. P. 27–36.
 28. Chari S. et al. Explanation Ontology: A Model of Explanations for User-Centered AI. 2020. P. 228–243.
 29. Xiong W. et al. Challenges of human—machine collaboration in risky decision-making // *Front. Eng. Manag.* 2022. Vol. 9. No. 1. P. 89–103.
 30. Dellermann D. et al. The future of Human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems // *Proc. Annu. Hawaii Int. Conf. Syst. Sci.* 2019. Vol. 2019-Janua. P. 274–283.
 31. Smirnov A. et al. Methodology for Multi-Aspect Ontology Development: Ontology for Decision Support Based on Human-Machine Collective Intelligence // *IEEE Access*. 2021. Vol. 9. P. 135167–135185.
 32. Sandkuhl K., Shilov N., Smirnov A. Facilitating Digital Transformation by Multi-Aspect Ontologies: Approach and Application Steps // *IFAC-PapersOnLine*. 2019. Vol. 52. No. 13. P. 1609–1614.
 33. Korzun D.G., Balandin S.I., Gurtov A. V. Deployment of Smart Spaces in Internet of Things: Overview of the Design Challenges. 2013. P. 48–59.

34. Roffia L. et al. A Semantic Publish-Subscribe Architecture for the Internet of Things // *IEEE Internet Things J.* 2016. Vol. 3. No. 6. P. 1274–1296.
35. The PROV Ontology [Electronic resource]. URL: <https://www.w3.org/TR/prov-o/> (accessed: 28.01.2020).

Смирнов Александр Викторович. Доктор технических наук, профессор. Главный научный сотрудник. Санкт-Петербургский Федеральный исследовательский центр РАН. Области исследований: системный анализ, интеллектуальная поддержка принятия решений, информационные технологии. E-mail: smir@iias.spb.su

Пonomarev Андрей Васильевич. Кандидат технических наук, доцент. Старший научный сотрудник. Санкт-Петербургский Федеральный исследовательский центр РАН. Области исследований: коллективный интеллект, машинное обучение, поддержка принятия решений. E-mail: ponomarev@iias.spb.su

Шилов Николай Германович. Кандидат технических наук, доцент. Старший научный сотрудник. Санкт-Петербургский Федеральный исследовательский центр РАН. Области исследований: конфигурирование сложных систем, машинное обучение. E-mail: nick@iias.spb.su

Левашова Татьяна Викторовна. Кандидат технических наук. Старший научный сотрудник. Санкт-Петербургский Федеральный исследовательский центр РАН. Области исследований: интеллектуальная поддержка принятия решений, информационные технологии, разработка онтологий, управление онтологиями, управление контекстом. E-mail: tatiana.levashova@iias.spb.su

Neuro-Symbolic Artificial Intelligence in Collaborative Decision Support Systems

A. V. Smirnov, A. V. Ponomarev, N. G. Shilov, T. V. Levashova

St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia

Abstract. The paper discusses the requirements for collaborative human-machine decision support systems and the problems that arise during their creation. The methods of neuro-symbolic artificial intelligence can help to resolve some of these problems. The analysis of modern results in the field of ontology-oriented neuro-symbolic artificial intelligence is carried out, primarily aimed at explaining neural network models using ontologies and using symbolic knowledge to improve the efficiency of neural network models. A conceptual model of a collaborative human-machine decision support system based on ontology-oriented neuro-symbolic intelligence is proposed.

Keywords: decision support systems, neuro-symbolic intelligence, collaborative systems.

DOI 10.14357/20718594220303

References

1. Averkin A.N. Obyasnimyj iskusstvennyj intellekt: itogi i perspektivy [Explainable artificial intelligence] // *Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte (IMMV-2021). Sbornik nauchnyh trudov X Mezhdunarodnoj nauchno-tekhnicheskoy konferencii [Integrated models and soft computing in artificial intelligence. Proceedings of the 10-th international conference]*. 2021. P. 153–174. (in Russian).
2. Demidovskiy A.V., Babkin E.A. Integrirovannye nejrosimvolicheskie sistemy podderzhki prinyatiya reshenij: problemy i perspektivy [Integrated neurosymbolical decision support systems: problems and perspectives] // *Biznes-informatika [Business informatics]*. 2021. Vol. 15. No. 3. P. 7–23 (in Russian).
3. Appelbaum S.H. Socio-technical systems theory: an intervention strategy for organizational development // *Manag. Decis.* 1997. Vol. 35.No. 6. P. 452–463.
4. Rafner J. et al. Deskillling, Upskilling, and Reskilling: a Case for Hybrid Intelligence // *MORALS + MACHINES*. 2021. P. 24–39.
5. Johnson M. et al. Coactive Design: Designing Support for Interdependence in Joint Activity // *J. Human-Robot Interact.* 2014. Vol. 3.No. 1. P. 43.
6. Bosch K. van den, Bronkhorst A. Human-AI Cooperation to Benefit Military Decision Making: STO-MP-IST-160. Soesterberg. 2018. 13 p.
7. Xiong W. et al. Challenges of human—machine collaboration in risky decision-making // *Front. Eng. Manag.* 2022. Vol. 9. No. 1. P. 89–103.

8. Utkin L.V. et al. Obzor metodov obyasneniya i interpretacii prinyatiya reshenij v intellektual'nyh sistemah diagnostiki onkologicheskikh zabolevanij [A review of methods for explaining and interpreting decisions of intelligent cancer diagnosis systems] // *Iskusstvennyj intellekt i prinyatie reshenij* [Artificial intelligence and decision-making]. 2020. No. 4. P. 55–65. (in Russian).
9. Cummings M.L. Automation and Accountability in Decision Support System Interface Design // *J. Technol. Stud.* 2006. Vol. 32. No. 1.
10. Kroon M.B.R., Van Kreveld D., Rabbie J.M. Group Versus Individual Decision Making: Effects of Accountability and Gender on Groupthink // *Small Gr. Res.* 1992. Vol. 23. No.4. P. 427–458.
11. *Mastering Data-Intensive Collaboration and Decision Making* / ed. Karacapilidis N. Cham: Springer International Publishing, 2014. Vol. 5.
12. Ribeiro de Sousa M., Leite J. Aligning Artificial Neural Networks and Ontologies towards Explainable AI // *Proceedings of the AAAI Conference on Artificial Intelligence.* 2021. Vol. 35. No.6. P. 4932–4940.
13. Confalonieri R. et al. Using ontologies to enhance human understandability of global post-hoc explanations of black-box models // *Artif. Intell.* Elsevier B.V. 2021. Vol. 296. P. 103471.
14. Gupta U., Chaudhury S. Deep transfer learning with ontology for image classification // 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). IEEE. 2015. P. 1–4.
15. Sun Y., Ghaffarzagdegan S. An Ontology-Aware Framework for Audio Event Classification // *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE. 2020. P. 321–325.
16. Phan N. et al. Ontology-based deep learning for human behavior prediction with explanations in health social networks // *Inf. Sci. (Ny).* 2017. Vol. 384. P. 298–313.
17. Kuang Z. et al. Integrating multi-level deep learning and concept ontology for large-scale visual recognition // *Pattern Recognit.* 2018. Vol. 78. P. 198–214.
18. Fong A.C.M., Hong G. Ontology-Powered Hybrid Extensional-Intensional Learning // *Proceedings of the 2019 International Conference on Information Technology and Computer Communications - ITCC 2019.* New York, New York, USA: ACM Press. 2019. P. 18–23.
19. Zha Y. et al. Ontology-Aware Deep Learning Enables Ultrafast, Accurate and Interpretable Source Tracking among Sub-Million Microbial Community Samples from Hundreds of Niches // *bioRxiv.* 2020.
20. Panigutti C., Perotti A., Pedreschi D. Doctor XAI: An ontology-based approach to black-box sequential data classification explanations // *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, Transparency.* 2020. P. 629–639.
21. Bourgeais V. et al. Deep GONet: self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data // *BMC Bioinformatics.* BioMed Central. 2021. Vol. 22. P. 1–24.
22. Voogd J. et al. Using Relational Concept Networks for Explainable Decision Support // 3rd IFIP Cross Domain Conference for Machine Learning and Knowledge Extraction, CD-MAKE 2019, LNC 11713. 2019. P. 78–93.
23. Abbass H.A. et al. Machine Education: Designing semantically ordered and ontologically guided modular neural networks // 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE. 2019. P. 948–955.
24. Bourguin G. et al. Towards Ontologically Explainable Classifiers // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2021. Vol. LNCS 12892. P. 472–484.
25. Neerinx M.A. et al. Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance // 15th International Conference on Engineering Psychology and Cognitive Ergonomics, EPCE 2018 Held as Part of HCI International 2018, LNAI 10906. Springer. 2018. P. 204–214.
26. Donadello I., Dragoni M. SeXAI: A Semantic Explainable Artificial Intelligence Framework // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2021. Vol. 12414. June. P. 51–66.
27. Zagorul'ko Yu. A. Sovremennye sredstva formalizacii semantiki oblastej znaniy na osnove ontologij [Modern means for formalizing the semantics of knowledge areas on the basis of ontologies] // *Informacionnye i matematicheskie tekhnologii v nauke i upravlenii* [Information and mathematical technologies in science and management]. 2018. Vol. 11. No. 3. P. 27–36. (in Russian).
28. Chari S. et al. Explanation Ontology: A Model of Explanations for User-Centered AI. 2020. P. 228–243.
29. Xiong W. et al. Challenges of human—machine collaboration in risky decision-making // *Front. Eng. Manag.* 2022. Vol. 9. No. 1. P. 89–103.
30. Dellermann D. et al. The future of Human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems // *Proc. Annu. Hawaii Int. Conf. Syst. Sci.* 2019. Vol. 2019-Janua. P. 274–283.
31. Smirnov A. et al. Methodology for Multi-Aspect Ontology Development: Ontology for Decision Support Based on Human-Machine Collective Intelligence // *IEEE Access.* 2021. Vol. 9. P. 135167–135185.
32. Sandkuhl K., Shilov N., Smirnov A. Facilitating Digital Transformation by Multi-Aspect Ontologies: Approach and Application Steps // *IFAC-PapersOnLine.* 2019. Vol. 52. No.13. P. 1609–1614.
33. Korzun D.G., Balandin S.I., Gurtov A. V. Deployment of Smart Spaces in Internet of Things: Overview of the Design Challenges. 2013. P. 48–59.
34. Roffia L. et al. A Semantic Publish-Subscribe Architecture for the Internet of Things // *IEEE Internet Things J.* 2016. Vol. 3. No.6. P. 1274–1296.
35. The PROV Ontology [Electronic resource]. URL: <https://www.w3.org/TR/prov-o/> (accessed: 28.01.2020).

Smirnov Alexander V. Doctor of technical sciences, professor. Chief Researcher. St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research areas: system analysis, intelligent decision support, information technologies, recommender systems. E-mail: smir@iiias.spb.su

Ponomarev Andrew V. Candidate of technical sciences, docent. Senior researcher. St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research areas: collective intelligence, applied machine learning, decision support. E-mail: ponomarev@iias.spb.su

Shilov Nikolay G. Candidate of technical sciences, docent. Senior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research areas: complex systems configuration, machine learning. E-mail: nick@iias.spb.su

Levashova Tatyana V. Candidate of technical sciences. Senior researcher. St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research areas: intelligent decision support, information technologies, ontology engineering, ontology management, context management. E-mail: tatiana.levashova@iias.spb.su