

Морской Ёж

Alexander Kuznetsov

17 февраля 2025 г.

1 Мотивация

Данная работа посвящена одной из возможных формализаций понятия субъективной вероятности (в том смысле, как это понимается логиками, начиная с F.P. Ramsey [17]) и связанными с этой формализацией теорией неполного или неточного знания. Субъективная вероятность обычно понимается как степень веры агента в то, что определенное событие произойдет, в отличие от статистической вероятности, понимаемой как частота, с которой происходит событие. Обычно ученые с негодованием отвергают неполное или неточное знание как заблуждение, равно как и события, которые происходят 1-2 раза и более не воспроизводятся. Однако, в реальности происходят единичные события и из полужнания или знання, запятнанного заблуждением, вполне возможно извлечь чистое, безошибочное знание, хотя существует не так много теорий этого процесса.

Возможным практическим применением предлагаемой теории являются автоматизированный сбор и анализ истинности информации, в том числе и в рамках информационного противоборства, и валидация вывода LLM и подобных нейросетевых механизмов с целью отбрасывания галлюцинаций (см., например, [14, 19]). Валидация осуществляется посредством отбрасывания вывода, оцениваемого как маловероятный (в смысле логической вероятности) в рамках заданной онтологии.

Рассмотрим несколько примеров, в которые демонстрируют проявления субъективной вероятности и демагогии при взаимодействии нескольких агентов, как минимум один из которых не полностью осведомлен о ситуации.

Модельный пример 1 (мы все – морские ежи, в каком-то смысле):

Пусть есть два лица – Простофиля и Некто. Некто говорит:

- (Посылка) Все люди – братья, так как они произошли от Адама и Евы*.

- (Посылка) Братья должны делить все поровну.
- (Посылка) Ты – человек и я – человек.
- (Закключение) Значит, ты должен со мной поделиться поровну.

* Можно – от морского ежа, обезьяны, древних эукариот, главное, чтоб эта посылка разделялась адресатом умозаключения, Простофилей.

Формально, с этим силлогизмом все хорошо, более того, все послылки могут разделяться адресатом. Фактически, если Простофиля предлагается поделиться средствами, к получению которых Некто не имеет отношения, то даже самый простой Простофиля почувствует подвох.

Модельный пример 2 (информационная война):

Двое шпионов обрабатывают информацию и одновременно стараются замедлить работу друг друга. Для обработки каждого блока информации B каждый из шпионов тратит T секунд. Шпионы последовательно обрабатывают блоки $A_1^j, A_2^j, \dots, A_t^j$, $j = 1$ или $j = 2$ соответственно. При этом, не все блоки информации нужны каждому шпиону для получения критически важных сведений, часть является бесполезными и просто тратящими ценное время.

Шпион j может, для экономии времени, отбросить блок A_t^j в момент времени, но это, с вероятностью $P(t)$, приведет к ошибочному выводу и необходимо будет опять обрабатывать ошибочно отброшенный блок.

Шпионы могут подбрасывать друг другу бесполезную информацию, чтобы замедлить противника.

Побеждает тот, кто получит правильный вывод в кратчайшее время.

Вопрос: как получить оценку риска при отбрасывании блоков информации с целью экономии времени?

Модельный пример 3 (детерминированная недетерминированность):

Коррумпированный судья рассматривает тяжбу между двумя сторонами, Жуликом и Простофилей. Простофиля подает все необходимые документы в поддержку своей позиции. Решение судьи для Простофили недетерминированно – все необходимые сигналы на вход поданы, а на внутреннее состояние судьи Простофиля, как и всякий честный гражданин, прямого влияния не имеет.

Жулик дает судье денег, и судья отказывает Простофиле в его требованиях, находя лазейку в законах. Для Жулика судья детерминирован.

Во всех примерах есть нечто общее: конкуренция двух лиц за ресурсы. Примеры 1 и 3 асимметричны, их участники в неравных положениях, пример 2 изображает симметричную конкуренцию.

Примеры 2 и 3 также близки тем, что для участников или участника процесса результат имеет вероятностный характер, тогда как глобально он полностью детерминирован.

Пример 1 можно также интерпретировать, как увеличение вероятности того, что Простофиля поделится с Некто ресурсами.

Модельный пример 4 (ложная вероятность):

Есть узкая коробка, на дне которой – два шара, черный справа и белый слева. Хитрец знает расположение шаров, а Простофиля – нет. Простофиля думает, что достанет белый шар с вероятностью $1/2$, тогда как Хитрец достает белый шар с вероятностью 1, если приходит к ящику первым.

Для построения формального описания ситуаций в модельных примерах 1 – 4 необходимы три вещи:

- ввести расстояние между сущностями в онтологиях, чтоб понимать, какие утверждения стоит принимать, а какие – отбросить, как слишком далекие от исследуемой области,
- градуировать понятие истинности, чтоб отделять просто истинные утверждения от истинных и полезных тому или иному агенту,
- для задач, включающих понятие вероятности, учитывать, что ни события, ни вероятностная мера не фиксированы, а зависят от агента, с которым происходят события и от времени наступления события.

В модельном примере 1, Простофиле достаточно взять в качестве «расстояния» между людьми TMRSA, т.е. возраста ближайшего общего предка. Очевидно, что отношения, возникшие более тысячи лет назад, вряд ли могут повлиять на распределение средств в настоящем и доля средств, причитающаяся самозваному «брату», как и ценность его утверждений обратно пропорциональна TMRSA. Таким образом, утверждению можно приписывать, помимо истинности, еще и ценность V , например

$$V = \min\{1, \frac{100}{TMRSA}\}$$

и делиться средствами пропорционально V .

В модельном примере 4, Простофиля думает, что множество событий – это

$$A = \{white, black\},$$

а вероятность достать белый шар $P\{white\} = 1/|A|$.

Однако, в реальности вероятность определяется как

$$P_t\{white\} = \begin{cases} 1/2, t = 1, \\ 0, t = 2, \end{cases}$$

если предположить, что в момент времени $t = 2$ к ящику подходит Хитрец и процесс утрачивает вероятностный характер. Таким образом, на самом деле, множество событий – это

$$A = \{(white, 1), (black, 1), (white, 2), (black, 2)\}$$

и речь для Простофили идет скорее о случайном процессе.

В модельном примере 2, каждый из шпионов должен определить онтологическое расстояние для анализируемых ими сущностей, и определить ценность V_t^j каждого блока A_t^j как величину, обратно пропорциональную расстоянию от блока A_t^j до множества M , внутри которого шпион ищет нужные ему сведения.

Можно далее предположить, что вероятность ошибочно отбросить блок A_t^j прямо пропорциональна его ценности и определяется, например, как

$$P_t^j = V_t^j / \max_{j,t} V_t^j.$$

Позже мы вернемся к теоретико-игровой интерпретации этой задачи.

В модельном примере 3 взаимоотношения Судьи и Простофили описываются недетерминированным конечным автоматом $\mathcal{A}$. Однако, глобальные отношения Судьи, Простофили и Жулика описываются детерминированным конечным автоматом, подавтоматом которого является $\mathcal{A}$. В терминах случайных процессов, можно сказать, что эти взаимоотношения описываются случайным процессом $\xi(t, a(t))$, ξ принимает значения «Простофиля выиграл», «Простофиля проиграл», $t \geq 0$, причем $a(t)$ – это также случайный процесс,

$$a \in \{\text{«Жулик дал взятку»}, \text{«Жулик не дал взятку»}\}$$

таким что если в момент $t = t_0$ оказывается, что

$$a(t) = \text{«Жулик дал взятку»},$$

то для $t \geq t_0$ случайный процесс $\xi(t, a(t))$ превращается детерминированный с

$$\xi(t, a(t)) = \text{«Простофиля проиграл»}.$$

2 Обзор

Существует несколько подходов для формализации описанной выше «субъективной вероятности» и «полезной» и «бесполезной» истин.

Полезность утверждения может быть выражена через расстояние между концепциями в онтологии (в том смысле, как онтология понимается математиками [9]).

В статье [16] представлен подход DIS-C, который является новым методом оценки концептуального расстояния между концепциями в онтологии. DIS-C основан на графе в том смысле, что при вычислении веса отношений между концепциями учитывается вся топология онтологии. Методология состоит из двух основных шагов. Во-первых, чтобы воспользоваться предыдущими знаниями, эксперт в области онтологии назначает начальные значения веса каждому из отношений в онтологии. Затем автоматический метод вычисления концептуальных отношений уточняет веса, назначенные каждому отношению, до достижения стабильного состояния. Мы вводим метрику, называемую общностью, которая определяется для оценки доступности каждого понятия, рассматривая онтологию как сильно связанный граф. В отличие от большинства предыдущих подходов, алгоритм DIS-C вычисляет сходство между концепциями в онтологиях, которые не обязательно представлены в иерархической или таксономической структуре. Таким образом, DIS-C способен включать широкий спектр отношений между концепциями, такими как меронимия, антонимия, функциональность и причинность.

Более того, можно сравнивать и онтологии целиком [7]. Расстояния между онтологиями полезны для поиска, сопоставления или визуализации онтологий. Мы изучаем различные расстояния, которые могут быть определены в онтологиях, и предоставляем их в плагине NeOn toolkit, OntoSim, который представляет собой библиотеку расстояний, которые можно использовать для реконтекстуализации.

Одна из важнейших концепций при сравнении онтологий – это Semantic similarity [11]. В конечном счете, семантическое сходство выливается в сходство графов, если онтология представлена графом, или в семантические меры на основе корпуса, которые позволяют сравнивать единицы языка из анализа неструктурированных или полуструктурированных текстов. Они чаще всего используются для сравнения слов, предложений или текстов основанных на методах обработки естественного языка, которые чаще всего полагаются только на статистический анализ использования слов в текстах, например, основанных на анализе (совместного) употребления слов и лингвистических контекстов, в которых они встречаются.

Наиболее систематическое исследование логической вероятности принадлежит Р. Карнапу [12]. Его формулировка логической вероятности [3] начинается с построения формального языка. Он рассматривает класс языков, состоящий из конечного числа логически независимых монадических предикатов (naming properties), применяемых к счетному множеству отдельных констант (naming individuals) или переменных, а также обычных логических связок. Самые сильные (последовательные) утверждения, которые можно сделать на данном языке, описывают универсум настолько подробно, насколько позволяет выразительная сила языка. Они представляют собой конъюнкции полных описаний каждого индивидуума, причем каждое описание само по себе является конъюнкцией, содержащей ровно одно вхождение (с отрицанием или без него) каждого предиката языка. Эти самые сильные утверждения называются описаниями состояний.

Любая вероятностная мера по описаниям состояний автоматически расширяется до меры по всем предложениям, поскольку каждое предложение эквивалентно дизъюнкции описаний состояний. Можно задать вероятностную меру для описания состояний так, чтобы состояния, которые бы соответствовали нескольким возможным состояниям универсума, имели бы меньшую вероятность, чем состояния, соответствующие только одному состоянию универсума.

Прежде чем переходить к описанию различных вероятностных логических систем, укажем, что сравнивать онтологии возможно не только сравнивая топологию онтологий, но и атрибуты объектов онтологий, используя инструментарий грубых множеств [15]. В данной теории рассматриваются множества объектов, имеющих атрибуты. Эти множества разделены на классы эквивалентности по отношению неразличимости (P -indiscernibility relation), отождествляющему объекты с совпадающими в каком-то смысле атрибутами (например, таких, для которых совпадают значения заданной на множестве атрибутов функции). Верхней P -аппроксимацией множества называется объединение всех классов эквивалентности, имеющих непустое пересечение с данным множеством, нижней P -аппроксимацией множества – объединение всех классов эквивалентности, входящих в данное множество. Любое множество может быть приближено грубым множеством – парой верхней и нижней P -аппроксимации множества.

В работе [8] показано, что хотя логика обеспечивает привлекательный формализм для представления знаний и рассуждений, она не лишена недостатков. Существует множество аспектов человеческого рассуждения, которые неадекватно обрабатываются чистой логикой. (Конечно, есть также множество аспектов, которые мы предпочли бы избегать.)

Например, предположим, что человеку говорят, что «все инопланетяне синие». Однако позже человек также узнает тот факт, что «инопланетяне типа клингон не синие». Хотя этот новый факт противоречит исходному утверждению, у человека не возникнет никаких трудностей с разрешением конфликта. Большинство людей будут рассматривать инопланетян типа клингон как особый случай, и в противном случае все остальные инопланетяне будут считаться синими. Однако, компьютерная система автоматического вывода не справится с таким умозаключением. Теперь мы опишем простой подход к расширению логики для обработки несоответствий и немонотонных шаблонов вывода. Идея состоит в том, чтобы присвоить вес каждой логической формуле, которая интерпретируется как «стоимость» нарушения формулы. Ниже мы представляем эту идею и обсуждаем основные подходы к проблемам рассуждения и обучения.

Системы такого типа, можно сказать - «с нечетким выводом» следует отличать от систем нечеткой логики, в которых градуировано само понятие истинности, а не логические формулы, см. статью [6], в которой содержится структурированный вводный обзор ландшафта взвешенных логик (в общем смысле), которые можно найти в литературе по искусственному интеллекту, подчеркивая их фундаментальные различия и области их применения.

Идея применения теории вероятности в логике довольно старая и есть еще в работе [4]. Кокс (см. Cox's theorem) хотел, чтобы его система удовлетворяла следующим условиям:

- Делимость и сравнимость – Правдоподобность предложения является действительным числом и зависит от информации, которую мы имеем относительно предложения.
- Здравый смысл – Правдоподобия должны разумно варьироваться в зависимости от оценки правдоподобий в модели.
- Согласованность – если правдоподобность предложения может быть выведена многими способами, все результаты должны быть равны.

Родственным данному подходу является подход с evidence theory или Dempster-Shafer theory (DST) [5]. Теория Демпстера–Шейфера является обобщением байесовской теории субъективной вероятности. Функции убеждения основывают степени убеждения (или уверенности, или доверия) для одного вопроса на субъективных вероятностях для связанного вопроса. Сами степени убеждения могут иметь или не иметь математические свойства вероятностей; насколько они различаются, зависит от того, насколько тесно связаны два вопроса.

Теория Демпстера–Шейфера основана на двух идеях: получение степеней убеждения для одного вопроса из субъективных вероятностей для связанного вопроса и правило Демпстера для объединения таких степеней убеждения, когда они основаны на независимых элементах доказательств. По сути, степень убеждения в предложении зависит в первую очередь от количества ответов (на связанные вопросы), содержащих предложение, и субъективной вероятности каждого ответа. Также свой вклад вносят правила комбинирования, которые отражают общие предположения относительно данных.

3 Описание системы

3.1 Онтология и ее отражения

Пусть задано множество агентов $Ag = \{ag_1, \dots, ag_m\}$ и онтология $\mathbf{O}$. Для простоты будем считать, что онтологии соответствует граф онтологии

$$\Gamma = (\mathcal{V}, \mathcal{E}, \Lambda_{\mathcal{V}}, \Lambda_{\mathcal{E}}, I, \varphi_{\mathcal{V}}, \varphi_{\mathcal{E}}),$$

вершины $\mathcal{V}$ которого соответствуют экземплярам и понятиям, а множество ребер с подлежащим множеством $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ – отношениям онтологии. Также задана функция индидентности

$$I : \mathcal{V} \times \mathcal{V} \rightarrow \{0, 1\},$$

функция атрибутов вершин $\varphi_{\mathcal{V}} : \mathcal{V} \rightarrow A \subseteq \mathbb{R}^p$, функция атрибутов ребер $\varphi_{\mathcal{E}} : \mathcal{E} \rightarrow B \subseteq \mathbb{R}^q$, $p > 0$, $q > 0$, предполагается, что все атрибуты закодированы векторами вещественных чисел. Символ 0 обозначает «Ложь», символ 1 обозначает «Истина».

Для каждого агента определено отражение онтологии – множество известной ему части онтологии $\mathbf{O}_{ag}$ и соответствующего ей графа онтологии

$$\Gamma_{ag} = (\mathcal{V}_{ag}, \mathcal{E}_{ag}, \Lambda_{\mathcal{V}_{ag}}, \Lambda_{\mathcal{E}_{ag}}, I, \varphi_{\mathcal{V}_{ag}}, \varphi_{\mathcal{E}_{ag}}),$$

где $\mathcal{V}_{ag}$ – множество вершин Γ_{ag} , $\mathcal{E}_{ag}$ – множество ребер Γ_{ag} , I_{ag} – функция индидентности Γ_{ag} , функция атрибутов вершин $\varphi_{\mathcal{V}_{ag}} : \mathcal{V}_{ag} \rightarrow A_{ag} \subseteq \mathbb{R}^p$, функция атрибутов ребер $\varphi_{\mathcal{E}_{ag}} : \mathcal{E}_{ag} \rightarrow B_{ag} \subseteq \mathbb{R}^q$.

Обратим внимание, что знания агента об онтологии могут не соответствовать самой онтологии и Γ_{ag} может не являться подграфом Γ .

В онтологию должны ходить правила вывода $\mathbf{R} = \{R_j\}$ и соответствующий им граф $\Gamma_{\mathbf{R}} \subset \Gamma$ и они должны быть известны каждому агенту $ag \in Ag$, т.е. $\Gamma_{\mathbf{R}} \subset \Gamma_{ag}$.

Пусть для каждого агента $ag \in Ag$ задано множество $P_{\text{for}_{ag}}$ возможных утверждений о сущностях онтологии $\mathbf{O}$. Эти утверждения можно поделить на ложные, истинные, бессмысленные 1-го типа и бессмысленные 2-го типа.

Истинны лишь умозаклучения о свойствах вершин, ребер и их инцидентности, совпадающие с теми, что есть в графе Γ на самом деле. Также возможны бессмысленные утверждения о ребрах и вершинах, которых в Γ нет.

Истинное или ложное утверждения P_0 о сущностях $\mathbf{O}$ включает в себя утверждения вида

$$P_{00}(V_1, V_2) : I_{ag}(V_1, V_2), V_1, V_2 \in \mathcal{V}, \quad (1)$$

$$P_{01}(V) : \varphi_{\mathcal{V}_{ag}}(V) = \alpha \in A \subseteq \mathbb{R}^p, \quad (2)$$

$$P_{02}(E) : \varphi_{\mathcal{E}_{ag}}(E) = \beta \in B \subseteq \mathbb{R}^q, \quad (3)$$

а также производные от вышеперечисленных логические формулы и выводимые из них с помощью правил вывода $\mathbf{R}$ формулы.

Бессмысленные утверждения первого типа P_1 включают в себя утверждения о вершинах, ребрах и атрибутах, которых в графе Γ нет, но, при этом, граф Γ может быть вложен в некоторый граф-расширение $\tilde{\Gamma}_{ag}$, в котором такие вершины, ребра и атрибуты возможны, и, соответственно, для $\tilde{\Gamma}_{ag}$ эти утверждения истинны или ложны:

1. $I_{ag}(V_1, V_2), V_1, V_2 \in \mathcal{V}_{ag} \setminus \mathcal{V}$,
2. $\varphi_{\mathcal{V}_{ag}}(V) = \alpha \in A_{ag} \setminus A, V \in \mathcal{V}$,
3. $\varphi_{\mathcal{E}_{ag}}(E) = \beta \in B_{ag} \setminus B, E \in \mathcal{E}$,
4. $\varphi_{\mathcal{V}_{ag}}(V) = \alpha \in A, V \in \mathcal{V}_{ag} \setminus \mathcal{V}$,
5. $\varphi_{\mathcal{E}_{ag}}(E) = \beta \in B, E \in \mathcal{E}_{ag} \setminus \mathcal{E}$,
6. $\varphi_{\mathcal{V}_{ag}}(V) = \alpha \in A_{ag} \setminus A, V \in \mathcal{V}_{ag} \setminus \mathcal{V}$,
7. $\varphi_{\mathcal{E}_{ag}}(E) = \beta \in B_{ag} \setminus B, E \in \mathcal{E}_{ag} \setminus \mathcal{E}$,

а также производные от вышеперечисленных логические формулы и выводимые из них с помощью правил вывода $\mathbf{R}$ формулы.

Бессмысленные утверждения второго типа P_2 включают в себя утверждения, которые совершены в отношении объектов, которые не представимы как вершины, ребра и атрибуты Γ и для которых невозможно (в

рамках ограничений, которые мы изначально устанавливаем при формализации онтологии) построить граф $\tilde{\Gamma}_{ag} \supset \Gamma$, для которого эти утверждения будут истинными или ложными. Например, для подхода, в котором вообще все возможные в мире атрибуты уже представлены в виде чисел, это будет утверждение типа

$$I_{ag}(\text{Кря-кря, Запупыринка}) = 0$$

и, тем более, некорректно составленные формулы типа

$$I_{ag}(\text{Кря-кря, } \sqrt{2}/) = (.$$

Мы будем далее считать, что агенты не совершают бессмысленных утверждений второго типа, хотя, в реальности, они вполне распространены при неформальных рассуждениях и не являются столь бесполезными, как может казаться, так как каждое бессмысленное утверждение второго типа P порождает семейство бессмысленных утверждений второго типа или даже истинных и ложных утверждений, близких к P в некоторой метрике формул, сравниваемых как строки или как деревья формул (см., например, [2]).

Очевидно, что чем меньше мощности $|\mathcal{V}_{ag} \setminus \mathcal{V}|, |\mathcal{E}_{ag} \setminus \mathcal{E}|, |A_{ag} \setminus A|, |B_{ag} \setminus B|$, тем меньше вероятность агента совершить бессмысленное утверждение 1-го типа P_1 , которое, в общем случае, бесполезно. Однако, иногда лучше разрешать агенту фантазировать о свойствах онтологии, чем жестко ограничивать возможные атрибуты, вершины и ребра, так как в противном случае мы можем уже получить невозможность описать онтологию в доступных нам терминах и бессмысленные утверждения 2-го типа.

Истинные или ложные утверждения P_0 порождают граф Γ_{ag}^0 , причем для истинных утверждений подтипа 1 и их конъюнкций порожденный граф Γ_{ag}^0 может быть вложен как подграф в граф Γ , возможно, игнорируя разметку вершин и ребер. Бессмысленные утверждения типа 1 P_1 также порождают граф Γ_{ag}^1 , который может быть вложен как подграф в граф-расширение $\tilde{\Gamma}_{ag}$.

Пусть агент ag высказывает только истинные утверждения о $\Gamma \setminus \Gamma_{ag}$, тогда как о $\Gamma \setminus \Gamma_{ag}$ он случайным образом высказывает как истинные, так и ложные утверждения. Вероятность совершения агентом ложного утверждения о графе Γ прямо пропорциональна мощности множеств вершин, ребер и их атрибутов в графе $\Gamma \setminus \Gamma_{ag}$.

Сами атомарные высказывания можно поделить как (1) – (3). Однако, так как высказывание $P_{00}(V_1, V_2)$ – это, в сущности, высказывание «Объект V_1 имеет отношение к объекту V_2 », что бессодержательно, то целесообразно рассматривать вместо $P_{00}(V_1, V_2)$ как атомарное

$P'_{00}(V_1, V_2, \beta) = P_{00}(V_1, V_2) \wedge P_{02}(V_1, V_2) = \beta$, т.е. «Объект V_1 имеет отношение типа β к объекту V_2 ».

Пусть

$$\mathcal{P}_{ag}^0 = \{P'_{00}(V_1, V_2, \beta), P_{01}(V), P_{02}(E) | V \in \mathcal{V}, \beta \in B\}$$

– множество всех возможных высказываний $ag \in Ag$ вышеперечисленных типов. Для любого $P \in \mathcal{P}_{ag}$ задана вероятность $\mathbb{P}\{P = \mathbf{1}\} = \pi(P)$.

Любое высказывание вида $f(P_1, \dots, P_q)$, f – булева функция q переменных, $P_i \in \mathcal{P}_{ag}$, $i = \overline{1, q}$ о графе Γ может быть представлена как совершенная дизъюнктивная нормальная форма:

$$f(P_1, \dots, P_q) = \bigvee_{f(\sigma_1, \dots, \sigma_q) = \mathbf{1}} \bigwedge_{i=1}^q P_i^{\sigma_i},$$

где

$$x^\sigma = \begin{cases} x, & \sigma = 1; \\ \bar{x}, & \sigma = 0. \end{cases}$$

По данной форме можно рассчитать логические вероятности соответствующих конъюнкций и дизъюнкций аналогично [3] и другие подобные меры.

Обозначим $\bar{\sigma} = (\sigma_1, \dots, \sigma_q)$, $\text{Truth}(P) = 1$ тогда и только тогда, когда утверждение $P = \mathbf{1}$ и определим следующую меру «правильности» утверждения

$$\text{Correct}(f(P_1, \dots, P_q)) = \left(\sum_{0 \leq |\bar{\sigma}| \leq 2^q - 1} f(\bar{\sigma}) \right)^{-1} \bigvee_{f(\bar{\sigma}) = \mathbf{1}} \frac{1}{q} \bigwedge_{i=1}^q \text{Truth}(P_i^{\sigma_i}).$$

Для истинных утверждений $f(P_1, \dots, P_q)$ верно, что

$$\text{Correct}(f(P_1, \dots, P_q)) = 1,$$

для ложных $\text{Correct}(f(P_1, \dots, P_q)) = 0$. Чем длиннее утверждение и чем больше в нем истинных членов $P_i^{\sigma_i}$, тем $\text{Correct}(f(P_1, \dots, P_q))$ ближе к 1, и наоборот, чем короче утверждение и чем меньше в нем истинных членов $P_i^{\sigma_i}$, тем $\text{Correct}(f(P_1, \dots, P_q))$ ближе к 0.

Наша мера «правильности» указывает, например, насколько много полезной информации об онтологии содержится в утверждении агента. Далее мы сформулируем подобную меру на языке графов.

3.2 Formal definition of the graph dissimilarity

Приведем определение метрики несходства графов $\rho : 2^{\tilde{\Gamma}_{ag}} \times 2^{\tilde{\Gamma}_{ag}} \rightarrow [0, 1]$, определенной в [18].

Let (V, ρ) , (W, ξ) are metric spaces. We define $\mathcal{G}$ as the set of simple graphs $\Gamma = (V', E', W', I, \eta)$ with weighted edges, $V' \subseteq V$, $W' \subseteq W$ are finite sets, $I : (V')^2 \rightarrow \{\text{True}, \text{False}\}$ is an incidence function, $\eta : (V')^2 \rightarrow W'$ is the weight function, and

$$E' = \{(v_1, v_2) | v_1, v_2 \in V' \wedge v_1 I v_2\}.$$

Let $\mathcal{G}^*$ is the set of bijections $\chi : V \rightarrow V$ and $\mathcal{G}_{adm}^* \subseteq \mathcal{G}^*$.

Let us choose $\chi \in \mathcal{G}_{adm}^*$ and $\Gamma_i \in \mathcal{G}$, $i = 0, 1$. The map χ can be an isomorphism between some subgraphs of Γ_0 and Γ_1 . Let $\Gamma'_i \subset \Gamma_i$ are the maximal subgraphs that $\chi : \Gamma'_0 \rightarrow \Gamma'_1$ is the graph's isomorphism, $\Gamma'_i = (V''_i, E''_i, W''_i I_i, \eta_i)$. The “poorness” of the map χ as an isomorphism of Γ_i can be formally expressed as

$$q(\chi; \Gamma_0, \Gamma_1) = F(b(\chi; \Gamma'_0, \Gamma'_1), d_0(\chi; \Gamma'_0, \Gamma'_1), d_1(\chi; \Gamma'_0, \Gamma'_1)),$$

where F is the non-negative, monotonically increasing in each variable function, $F(x, y, z) = 0$ iff $x = y = z = 0$, $d_i = |V'_i \setminus V''_i|$, b is the dissimilarity between Γ'_0 and Γ'_1 . Let also β is the non-negative, monotonically increasing in each variable function, $W(x, y) = 0$ iff $x = y = 0$, and

$$\mathbf{v} = \{\rho(v, \chi(v)) | v \in V''_0\}, \quad \mathbf{e} = \{\xi(\eta(v_1, v_2), \eta(\chi(v_1), \chi(v_2))) | v_1, v_2 \in V''_0, v_1 I_0 v_2\}, \quad (4)$$

$$\mathbf{v} = \{0\}, V''_0 = \emptyset, \quad \mathbf{e} = \{0\}, E''_0 = \emptyset. \quad (5)$$

In this case, we can define

$$b(\chi; \Gamma'_0, \Gamma'_1) = \beta(\mathbf{v}, \mathbf{e}).$$

Further, we will assume that $\beta(x, y)$ is a norm of the vector (x, y) . For example, it can be Chebyshev norm

$$\beta(\mathbf{v}, \mathbf{e}) = \|(\mathbf{v}, \mathbf{e})\|_\infty = \max\{\max \mathbf{v}, \max \mathbf{e}\},$$

taxicab distance's norm, Euclidian norm, and so forth.

Definition 1 Define the dissimilarity distance between Γ_0 and Γ_1 as the following

$$diss(\Gamma_0, \Gamma_1) = \min_{\chi \in \mathcal{G}_{adm}^*} q(\chi; \Gamma_0, \Gamma_1).$$

3.3 Достоверность и полезность информации

Пусть истинное, ложное или бессмысленное типа 1 утверждение порождает некоторый граф Γ_{ag} . Можно определить объективную «достоверность» утверждения (и порожденного им графа) как сходство Γ_{ag} и Γ , определяемое с помощью той или иной метрики несходства графов $\rho : 2^{\tilde{\Gamma}_{ag}} \times 2^{\tilde{\Gamma}_{ag}} \rightarrow [0, 1]$, например, определенной в [18]. Такая мера будет связана с «правильностью» утверждения из подраздела 3.1.

Определим $\psi_V : V_0 \times V_1 \rightarrow [0, 1]$ как функцию несходства вершин в графах Γ_0, Γ_1 , со множествами вершин V_0, V_1 и ребер E_0, E_1 , функциями инцидентности I_0, I_1 соответственно, $v_0 \in \Gamma_0$. Положим, что $\varphi_V(v_0, \chi(v_0)) = 1$ для вершины $v_0 \in \Gamma_0$, которая не соответствует ничему в Γ_1 , обозначим множество таких вершин как $V(\chi)$. Далее, $\varphi_V(\chi^{-1}(v_1), v_1) = 1$ для вершины $v_1 \in \Gamma_1$, которая не соответствует ничему в Γ_0 , обозначим множество таких вершин как $N_V(\chi)$. Для всех прочих вершин, которые входят в некоторый подграф Γ_0 , изоморфный (т.е. в котором смежные вершины соответствуют смежным вершинам) подграфу Γ_1 , $\varphi_V(v_0, \chi(v_0)) \in [0, 1]$, обозначим такое множество вершин изоморфных подграфов как $Q_V(\chi)$. Итоговую меру несходства определим как

$$\varphi_V(\chi) = \frac{\sum_{v_0 \in Q_V(\chi)} \psi_V(v_0, \chi(v_0)) + |M_V(\chi)| + |N_V(\chi)|}{|Q_V(\chi)| + |M_V(\chi)| + |N_V(\chi)|}.$$

Аналогично, определим меру несходства ребер. Пусть пары смежных вершин (v_0, v'_0) и (v_1, v'_1) соответствуют друг другу в графах Γ_0 и Γ_1 , $v_1 = \chi(v_0)$, $v'_1 = \chi(v'_0)$. При этом пусть (v_0, v'_0) соответствует множество $E_0(v_0, v'_0)$ ребер, (v_1, v'_1) соответствует множество $E_1(v_1, v'_1)$ ребер. Пусть задано множество $\mathcal{G}_E$ функций отождествления ребер графов $\chi_E : E_0(v_0, v'_0) \rightarrow E_1(v_1, v'_1)$.

Определим функцию несходства ребер $\psi_E : E_0 \times E_1 \rightarrow [0, 1]$. Если ребру $e_0 \in E_0(v_0, v'_0)$ не соответствует ничего в $E_1(v_1, v'_1)$, то положим $\psi_E(e_0, \chi_E(e_0)) = 1$, обозначим множество таких ребер как $M_E(\chi_E) \subseteq E_0(v_0, v'_0)$. Если ребру $e_1 \in E_1(v_1, v'_1)$ не соответствует ничего в $E_0(v_0, v'_0)$, то положим $\psi_E(\chi_E^{-1}(e_1), e_1) = 1$, обозначим множество таких ребер как $N_E(\chi_E) \subseteq E_1(v_1, v'_1)$. В прочих случаях $\psi_E(e_0, \chi_E(e_0)) \in [0, 1]$ и множество таких ребер со взаимным соответствием обозначим как $Q_E(\chi_E) \subseteq E_0(v_0, v'_0)$. Итоговую меру несходства всех ребер, соответствующих (v_0, v'_0) , определим как

$$\tilde{\varphi}_E(\chi_E; v_0, v'_0) = \frac{\sum_{e_0 \in Q_E(\chi_E)} \psi_E(e_0, \chi_E(e_0)) + |M_E(\chi_E)| + |N_E(\chi_E)|}{|Q_E(\chi_E)| + |M_E(\chi_E)| + |N_E(\chi_E)|}.$$

Обозначим как $S(\chi) = \{(v_0, v'_0) \in Q_V^2(\chi) | v_0 I_0 v'_0 = \mathbf{1}\}$ множество пар всех смежных вершин в подграфе Γ_0 , изоморфном подграфу $\chi(\Gamma_0) \subseteq \Gamma_1$. Итоговая мера несходства всех ребер графа определяется как

$$\varphi_E(\chi, \chi_E) = \frac{1}{|S(\chi)|} \sum_{(v_0, v'_0) \in S(\chi)} \tilde{\varphi}_E(\chi_E; v_0, v'_0).$$

Наконец, определим несходство графов Γ_0 и Γ_1 как

$$\text{diss}(\Gamma_0, \Gamma_1) = \min_{\chi \in \mathcal{G}_{adm}^*, \chi_E \in \mathcal{G}_E} \frac{\alpha_V \varphi_V(\chi) + \alpha_E \varphi_E(\chi, \chi_E)}{\alpha_V + \alpha_E},$$

$\alpha_1 > 0, \alpha_2 > 0$.

Дополнительно, чтобы показать, что утверждение о многих предметах, в котором есть ошибка лишь относительно нескольких предметов лучше, чем утверждение о немногих предметах, в котором есть ошибки о многих из них, можно модифицировать достоверность так, чтоб она соответствовала размеру графа.

Пусть на множестве $2^{\tilde{\Gamma}_{ag}}$ таким образом задано семейство отношений нестрогого порядка (субъективной «полезности») $\preceq_{ag}$. Будем говорить, что информация $\Gamma_1 \subseteq \tilde{\Gamma}_{ag}$ полезнее для агента $ag \in Ag$ информации $\Gamma_2 \subseteq \tilde{\Gamma}_{ag}$ если

$$\Gamma_2 \preceq_{ag} \Gamma_1.$$

Для упрощения, можно считать, что чем информация достовернее, тем она полезнее и это отношение порождено семейством функций полезности $u_{ag} : 2^{\tilde{\Gamma}_{ag}} \rightarrow \mathbb{R}_{\geq 0}$ и $\Gamma_2 \preceq_{ag} \Gamma_1$ если и только если $u_{ag}(\Gamma_2) \leq u_{ag}(\Gamma_1)$, $u_{ag}(\Gamma') = 1 - \text{diss}(\Gamma', \Gamma)$.

Можно задать и более точную меру типа энтропии, которая позволит работать даже с теми истинными высказываниями, которые сами по себе не задают никакого подграфа графа Γ (например, высказывания об атрибутах вершин и ребер графа). Пусть также для каждого $\Gamma_{ag} \subseteq \Gamma$ и $ag \in Ag$ задано понятие информированности. Если агент ag может совершить умозаключение $P_{ag}(\Gamma_{ag})$ истинное с вероятностью 1, то будем говорить, что агент информирован относительно Γ' . Каждому графу можно приписать количество информации агента относительно этого графа, например как

$$S_{ag}(\Gamma) = \sum_{\Gamma_{ag} \in \Gamma, P_{ag} \in Prop_{ag}} \mathbb{P}\{P_{ag}(\Gamma_{ag}) = 1\} \ln \mathbb{P}\{P_{ag}(\Gamma_{ag}) = 1\},$$

где $Prop_{ag}$ – множество всех возможных утверждений агента.

Все содержимое данного подраздела необходимо для того, чтоб формализовать понятия «есть зерно истины», «почти верное утверждение», «несмотря на ошибки, получена важная информация» и т.п.

3.4 Процесс накопления знаний

Предположим, что агент стремится приблизить граф онтологии-отражения Γ_{ag} к графу онтологии Γ .

Пусть задано внешнее время $t \in \mathbb{Z}_{\geq 0}$ и в каждый момент t агент ag анализирует некоторый подграф $\Gamma_{ag}^O(t) \subseteq \Gamma$ и должен совершить некоторое умозаключение $P_{ag}(t)$ об $\Gamma_{ag}^O(t)$, имеющее вероятностный характер, которое основывается на ранее совершенных умозаключениях

$$\mathcal{P} = \{P_{ag}(\tau) | \tau = \overline{1, t-1}\},$$

и на правилах вывода $\mathbf{R}$ и принять решение $D_{ag}(t)$ относительно $\Gamma_{ag}^O(t)$. Возможно, на принятие решения уйдет некоторое время $T(D_{ag}(t))$, пропорциональное мощности множеств ребер, вершин и их атрибутов в $\Gamma_{ag}(t)$. Результатом умозаключения должно быть новое отражение Γ_{ag} , соответствующее $P_{ag}(t)$ и $\Gamma_{ag}^O(t)$.

В практических приложениях, агентам редко требуется знание обо всем в мире, а скорее всего, они нуждаются в точном отражении лишь части графа Γ , соответствующей их интересам.

Обозначим через $\rho_{\mathbf{O}} : \mathbf{O} \times \mathbf{O} \rightarrow \mathbb{R}_{\geq 0}$ расстояние между объектами в онтологии $\mathbf{O}$. Будем считать, что это расстояние и онтология обладают таким свойством, что чем дальше на графе Γ в смысле какого-то расстояния на графе l_{Γ} находятся вершины, соответствующие объектам онтологии, тем больше между ними расстояние.

Пусть для $ag \in Ag$ задано множество интересующих его вещей и отношений Γ_{ag}^+ . Агент оценивает ценность информации $c(\Gamma_{ag}(t))$ по близости $\Gamma_{ag}(t)$ к Γ_{ag}^+ , например как

$$c(\Gamma_{ag}(t)) = \max_{\gamma_1 \in \Gamma_{ag}(t), \gamma_2 \in \Gamma_{ag}^+} l_{\Gamma}(\gamma_1, \gamma_2).$$

Однако, у информации (в том числе и у любого ложного и бессмысленного утверждения) есть и объективная ценность, определяемая отношением количества преобразований соответствующего этой информации графа, необходимых для превращения информации в истинное утверждение, и общего количества поддающихся преобразованию элементов графа (вершин, ребер и их атрибутов).

Итак, агент должен при итеративном построении онтологии-отражения руководствоваться двумя целями: минимизацией расходов времени на каждую итерацию-уточнение, что достигается неполным исследованием онтологии с учетом ценности информации, и минимизацией вероятности высказывания ложного или бессмысленного утверждения типа 1, что, напротив, зависит от точности соответствия онтологии-отражения агента онтологии.

Актуальность информации – это мера соответствия реальной и отражаемой онтологии в данный момент времени.

Каждый агент в момент времени t располагает своей неточной копией $\Gamma_{ag}(t)$ графа Γ . Чем ближе $\Gamma_{ag}(t)$ к Γ , тем выше информированность агента $S_{ag}(\Gamma)$.

3.4.1 Стохастическое и направленное улучшение отражений. Уточнение vs Расширение. Стратегии обучения

Смоделируем процесс улучшения графа-отражения онтологии агентом из предыдущего подраздела. Агент может рассматривать новые фрагменты графа, выбирая их случайно или выбирая их целенаправленно.

В первом случае, вершины $\Gamma_{ag}(t) \subseteq \Gamma$ выбираются случайно из вершин Γ каждый момент времени t . Во втором случае, вершины $\Gamma_{ag}(t) \subseteq \Gamma$ выбираются как наиболее близкие (в каком-то смысле, исходя из расстояния l_Γ на графе) к множеству интересующих агента вещей и отношений Γ_{ag}^+ или как наиболее близкие к уже известным вершинам $\Gamma_{ag}(t-1)$.

Далее, можно ввести вероятность построения ошибочного отражения $\Gamma_{ag}^O(t)$ в Γ_{ag} , а именно, ошибочного добавления или исключения вершин, ребер или атрибутов.

Агент может улучшать качество имеющегося отражения, повторно изучая $\Gamma_{ag}^O(t-1)$ или изучать новый подграф $\Gamma_{ag}^O(t)$. Иногда разумно предполагать, что при неоднократном изучении отражение и отражаемый подграф совпадут, т.е. если $\Gamma_{ag}^O(t_1) = \Gamma_{ag}^O(t_2) = \dots \Gamma_{ag}^O(t_w)$, $w > w_0 > 0$, то $\Gamma_{ag}(t_w) \supset \Gamma_{ag}^O(t_1)$.

Можно приписать вершинам и ребрам Γ разную притягательность для изучения агентом $ag \in Ag$ (не путать с полезностью из 3.3). Информация $\Gamma_1 \subset \Gamma$ может быть менее полезной, чем кажется, но агент склонен ее изучать, потому что она похожа на очень полезную информацию $\Gamma_2 \subset \Gamma$ в смысле метрики несходства графов ρ , при этом в смысле расстояния на графе l_Γ подграфы Γ_1 и Γ_2 расположены далеко друг от друга. Образно говоря, агент ищет рогатого зайца, и тратит много времени, чтоб поймать зайца, показавшегося ему рогатым, потому что в темноте агент принял уши за рога.

Можно предположить, что агент последовательно изучает путь на графе Γ , а также отклоняется к ближайшим к данному пути вершинам графа.

Наконец, агент может фантазировать и высказывать бессмысленные утверждения 1-го или 2-го типа, возможно – похожие (в смысле метрики несходства графов) на уже известную ему часть Γ , а потом искать ближайшие к ним в графе Γ или строить граф-расширение $\tilde{\Gamma}$ так, чтоб

высказанные им утверждения были осмысленными. Это соответствует смене научной парадигмы и интуитивным «гениальным озарениям», а не рутинному накоплению знаний.

4 Поиск в научных текстах с помощью вероятностных моделей

4.1 Онтология и ее отражение в научных текстах

Рассмотрим практический пример применения вышеописанной конструкции для классификации научных текстов и поиска в них различной информации на примере расширения онтологии на базе классификатора теории управления [21], описанной в работе [10]. Данная онтология встроена в Информационную систему анализа научной деятельности [13], а также используется для решения задач анализа научных и научно-технических текстов. Так, например, в работе [1] решается задача тематической сегментации научных публикаций. Это задача разделения неструктурированного текста на тематически связанные сегменты — такие, в которых речь идет об одном и том же. Применение данной онтологии при сегментации позволяет использовать больше информации о словах в тексте: помимо того чтобы основываться на классических алгоритмах, таких как со-осцилланс и семантических расстояниях, методы анализа онтологии позволяют применять расстояние между словами на графе, инкорпорируя тем самым фактологическую информацию из графа знаний в процесс принятия решений о разбиении текста на сегменты.

Перейдем к описанию расширенной онтологии научной деятельности. Пусть онтологии представляют из себя своеобразный «пирог» из верхней онтологии, усеченной онтологии и отраженной онтологии (рис. 1). Верхняя онтология — это идеальная онтология, описывающая все реально существующие объекты и все возможные связи между ними. Эту онтологию нецелесообразно или даже невозможно полностью выразить формальным языком. В нашем случае, это онтология всех возможных научных публикаций, которые были или будут когда-нибудь написаны.

Опишем верхнюю онтологию, представленную графом G . Множество вершин графа V состоит из

1. Терминов предметных областей научных направлений (например, «растение» V_{11}).
2. Терминов из сферы применения науки (например, «космос», «сельское хозяйство»).

3. Общенаучных (в том числе и метанаучных) терминов-классов (например, «предложение» V_{31} , «абзац» V_{32} , «подраздел» V_{33} , «раздел» V_{34}) и конкретных экземпляров этих классов, представленных строковыми литералами.
4. Терминов-классов математического аппарата (например, «граф» V_{41}) и конкретных экземпляров этих классов, представленных строковыми литералами.
5. Терминов, обозначающих научные поднаправления (подфакторы).
6. Терминов, обозначающих научные направления (факторы первого уровня, например «теория графов» V_{61} , «растениеводство» V_{62}).
7. Факторов нулевого уровня, т.е. символов пунктов 1–6 настоящего списка.
8. Правил вывода формальной (или неклассической) логики, переменные, константы и логические операции в которых закодированы с помощью специальных вершин Γ , а направление вывода определяется специальными ребрами, идущими от антецеденту к сукцеденту.

Опишем верхнюю онтологию, представленную графом Γ . Множество вершин графа $\mathcal{V}$ составляют:

1. термины из области математического аппарата — термины, которые используют для описания математических моделей, например:
 - «базис»,
 - «граф»,
 - «вероятность».
2. термины из предметной области — термины, которые используются для описания методов, например:
 - «система управления»,
 - «искусственный интеллект»,
 - «семантическая сеть».
3. термины из сферы применения — термины, которые используются для описания приложений, например:
 - «управление предприятием»,

- «технологические процессы»,
 - «приборостроение».
4. общенаучные термины — термины, которые могут использоваться агентами в разных научных направлениях, например:
- термин «множество» может использоваться математиками в значении «набор объектов» или может не быть термином, если используется в контексте, где не подразумевается специальное математическое значение: «множество ученых занимаются этой проблемой, но результат так и не был получен»;
 - термин «пространство» может использоваться математиками для описания некоторого пространства, которое представляет собой множество с определенной структурой и аксиоматикой, а может использоваться учеными, которые занимаются проблемами космоса;
 - термин «задача» может использоваться в любом научном направлении.
5. названия научных направлений (факторов второго уровня, подфакторов или подтем), например:
- «знания и онтологии»,
 - «электричество и магнетизм»,
 - «разностные уравнения».
6. названия научных направлений (факторов первого уровня, факторов или тем), например:
- «математическая логика»,
 - «алгебра и теория чисел»,
 - «авиация».
7. названия научных метанаправлений (факторов нулевого уровня, метафакторов или метатем):
- «общенаучные термины»,
 - «математический аппарат»,
 - «предметная область»,
 - «сфера применения».

8. неклассифицированные ключевые слова, выделенные авторами публикаций (примеры взяты из [20]):

- «теоретическая основа эффективного государственного управления»,
- «основные понятия»,
- «интеграция социальных и общественных наук».

Множества сущностей, соответствующих пп. 1–7 вышеприведенного списка могут очевидным образом пересекаться.

Считается, что каждому строковому литералу, соответствующему предложению научного текста, сопоставлено синтаксическое дерево, описывающее отношения между терминами-подстроками. Вершины V_1 и V_2 связаны ребрами различных типов, если:

1. Термин V_1 входит в определение V_2 ¹. Например, «автоматическая суммаризация» является «процессом» получения краткого содержания «документа» с помощью «компьютерной программы». Вхождение в определение может быть определено как вхождение подстроки, соответствующей V_1 в строку, соответствующую V_2 , так и как вхождение V_1 в синтаксическое дерево, соответствующее определению V_2 в позиции определяющего термина (а не связки или вспомогательного текста).
2. Термин V_1 является видом термина V_2 . Например, «автоматическое повторное включение» является видом «управляющего воздействия».
3. Термин V_1 входит в класс V_2 в соответствии с некоторым классификатором K (например, графом классификатора K_{CS} ².
 - Если V_2 — фактор первого уровня, входящий в классификатор K , то он будет называться фактором метафактора V_1 . Например:
 - «комбинаторика» — фактор первого уровня, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является фактором метафактора «математический аппарат», то есть «комбинаторика» входит в класс «математический аппарат»;

¹Ср. граф глоссария ИСАНД <https://isand.ipu.ru/thesaurus>

²Ср. классификатор ИСАНД <https://i-sand.ipu.ru/ema/classifier>

- «исследование операций» — фактор первого уровня, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является фактором метафактора «предметная область», то есть «исследование операций» входит в класс «предметная область».
 - Если V_2 — фактор второго уровня, входящий в классификатор K , то он будет называться подфактором фактора V_1 . Например:
 - «экзоскелет» — фактор второго уровня, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является подфактором фактора «мехатроника», то есть «экзоскелет» входит в класс «мехатроника»;
 - «теория массового обслуживания» — фактор второго уровня, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является подфактором фактора «исследование операций», то есть «теория массового обслуживания» входит в класс «исследование операций».
 - Если V_2 — термин, входящий в классификатор K , то он будет называться термином фактора V_1 . Например:
 - «фиксированный скафандр» — термин, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является термином подфактора «экзоскелет», то есть «фиксированный скафандр» входит в класс «экзоскелет»;
 - «боевая единица» — термин, входящий в классификатор K_{CS} , он согласно классификатору K_{CS} является термином подфактора «военное дело», то есть «боевая» входит в класс «военное дело».
4. Термин V_1 встречается в публикации V_2 (взвешенный граф встречаемости <https://isand.ipu.ru/graphs>). Для части графа онтологии, соответствующей терминам и тексту публикации, предлагается использовать вес близости терминов $\chi = \sigma + \sum_{i=1}^4 \theta_i$, где σ — значение параметра типа публикации, которая может являться тезисами, статьей, монографией или чем-то иным, а θ_i является бинарным значением параметра близости терминов в предложении, абзаце, подразделе и разделе соответственно. Рассмотрим следующие примеры.
- Пусть типам публикации соответствуют следующие численные значения: тезисы — 10, статья — 20, монография — 30, а зна-

чения параметра близости терминов единичны, то есть каждой размерности единицы текста (предложение, абзац, подраздел, раздел) соответствует значение 1. Тогда вес близости терминов, где два термина встречаются в монографии в одном подразделе, вычисляется следующим образом: $\chi = 30 + 1 + 1 = 32$.

- Пусть типам публикации соответствуют следующие численные значения: тезисы – 1, статья – 10, монография – 100, а значения параметра близости терминов следующие: встречаемость терминов в одном предложении – 10, в одном абзаце – 5, в одном подразделе – 3, в одном разделе – 1. Тогда вес близости терминов, где два термина встречаются в монографии в одном подразделе, вычисляется следующим образом: $\chi = 100 + 1 + 3 = 104$.

5. Термин V_1 входит в тот же фрагмент синтаксического дерева, что и термин V_2 (например, если в публикации присутствует составной термин «теория управления в организационных системах», то термины «теория», «управление», «организация» и «система» связаны ребром).
6. Вершина V_1 входит в тематический именованный кластер V_2 . Тематический именованный кластер – это множество статей и других элементов онтологии, подобранных по теме V_2 , такие кластера динамические и меняются в зависимости от времени.
7. Вершина V_1 входит в методологию V_2 . Например, если задача V_1 может формулироваться в виде модели V_2 или задача V_1 может быть решена алгоритмом V_2 или алгоритм V_1 применялся к данным V_2 , тогда вершины V_1 и V_2 , являются смежными, а соответствующие ребра имеют типы «формулируется в виде модели» β_1 , «решается» β_2 , «применялся к данным» β_3).
8. Вершина V_1 входит в целеполагание V_2 (если для цели публикации V_2 решается задача V_1 , то V_1 смежна с V_2).
9. Вершины V_1 и V_2 типа «фактор» применяются в публикации V_3 . Если в тексте публикации встретились термины из классификатора, то ставится связь между подфакторами из разных факторов первого уровня и факторами первого уровня, которые содержат эти термины. Например, если в публикации встретились термины V_1' «граф» и $V_2' =$ «растение», то их факторы первого уровня V_1 «Теория графов» и V_2 «Сельское хозяйство» будут связаны ребром, то

есть будет справедливо высказывание «теория графов применяется в сельском хозяйстве»).

Выше приведены девять типов ребер, однако, типология ребер может быть расширена за счет введения новых типов связей.

Для конкретной задачи используется усеченная онтология, которую далее будем называть онтологией, содержащая вершины верхней онтологии. Типы ребер выбираются из верхней онтологии согласно значимости того смысла, который вкладывается в значение типа ребра в зависимости от решаемой задачи поиска. Рассмотрим примеры истинных и ложных утверждений, которые могут быть получены на онтологии научной деятельности.

Пример 1. Введем онтологию O_1 следующим образом. Пусть вершинами онтологии будут все возможные термины V_i , где $i \in N$, которые имеют определения. Ребра онтологии O_1 определим как ребра первого типа, то есть если вершина V_1 входит в определение вершины V_2 , то будем считать, что вершины V_1 и V_2 являются соседями. Пусть вершины со значениями «состояние», «аварийный отказ», «неполные измерения» и «потребность» являются вершинами онтологии O_1 и определены следующим образом:

- состояние — набор значений переменных параметров объекта;
- потребность — состояние индивида, создаваемое испытываемой им нуждой, выступающее источником активности;
- неполные измерения — измерения, не охватывающие все параметры, описывающие состояние исследуемого объекта;
- аварийный отказ — переход объекта из работоспособного состояния в неработоспособное.

Тогда для онтологии O_1 будут **истинны** следующие высказывания:

- потребность — набор значений переменных параметров индивида, создаваемое испытываемой им нуждой, выступающее источником активности;
- неполные измерения — измерения, не охватывающие все параметры, описывающие набор значений переменных параметров объекта;
- аварийный отказ — переход объекта из работоспособного набора значений переменных параметров в неработоспособное.

Заметим, что последнее высказывание является некорректным, поскольку в онтологии $\mathbf{O}_1$ не определены термины «работоспособное состояние» и «неработоспособное состояние». Таким образом, данное высказывание **бессмысленно**, но это можно исправить, если добавить в онтологию термины «работоспособное состояние» и «неработоспособное состояние» и определить их. Следующие высказывания для онтологии $\mathbf{O}_1$ будут **ложными**:

- состояние — переход объекта из работоспособного состояния в неработоспособное;
- потребность — измерения, не охватывающие все параметры, выступающие источником активности;
- аварийный отказ — переход объекта из работоспособного набора значений переменных параметров индивида в неработоспособное.

Пример 2. Введем онтологию $\mathbf{O}_2$ следующим образом. Пусть вершинами V_i онтологии $\mathbf{O}_2$, где $i, j \in N$, будут термины, факторы и подфакторы из предметной области из сферы применения, а также сами метафакторы «предметная область» и «сфера применения». Будем считать, что вершины онтологии $\mathbf{O}_2$ входят в классификатор K_{CS} . Ребра онтологии $\mathbf{O}_2$ определим как ребра третьего типа, то есть если вершина V_1 входит в класс вершины V_2 , то будем считать, что вершины V_1 и V_2 являются соседями. Пусть вершины со значениями «энергетика», «электроэнергетика», «атомная энергетика», «авиация», «военное дело», «космос», «авиация», «военное дело», «космос», «авиация», «военное дело», «космос» и «потребность» являются вершинами онтологии $\mathbf{O}_2$ и связаны следующим образом в соответствии с классификатором K_{CS} :

- «энергетика» входит в класс «сфера применения»;
- «атомная энергетика» и «электроэнергетика» входят в класс «энергетика»;
- «модальный анализ» и «прогноз нагрузок» входят в класс «электроэнергетика»;
- «регулятор давления пара» и «тяжеловодные реакторы» входят в класс «атомная энергетика»;
- «прикладная лингвистика» входит в класс «предметная область»;
- «распознавание и синтез речи» входит в класс «прикладная лингвистика»;

- «интонация», «акустические признаки» и «шумоочистка» входят в класс «распознавание и синтез речи».

Тогда для онтологии O_2 будут **истинны** следующие высказывания:

- «электроэнергетика» входит в класс «сфера применения»;
- «интонация» входит в класс «предметная область»;
- «прикладная лингвистика» не входит в класс «сфера применения».

Следующие высказывания для онтологии O_1 будут **ложными**:

- «шумоочистка» входит в класс «сфера применения»;
- «интонация» входит в класс «атомная энергетика»;
- «энергетика» не входит в класс «сфера применения».

4.2 Процесс накопления научных знаний

Для определения процесса накопления знаний необходимы субъекты, которые будут частью этого процесса, которые меняются во времени с точки зрения системы.

Ученые образуют множество агентов Ag , которые в своих исследованиях порождают граф отражения онтологии. Взаимодействие между учеными может обогащать этот граф. Граф является динамическим: какие-то вершины и связи добавляются или удаляются в зависимости от обновления информации или добавления новых статей в систему, которое занимает время и грузит ресурсы. Таким образом есть некая актуальность информации, которой соответствует вероятность. Если информация, содержащая термин, всегда актуальна, соответствующая вероятность совершения утверждения с этим термином является 1. Онтология содержит базовый глоссарий - выделенные термины, которые мы всегда считаем актуальными. И добавочный глоссарий, который содержит новые термины, которые могут терять свою актуальность.

Онтология должна в качестве вывода давать не только сам вывод, но и вероятность его адекватности.

4.3 Автоматизированный анализ научных текстов

При автоматизированном анализе текстов с целью извлечения из него систематизированной информации, например, с помощью LLM, возможны т.н. галлюцинации.

Пример галлюцинации в ИС 1. Автору приписываются статьи, которые он никогда не писал.

Пример галлюцинации в ИС 2. Придуман несуществующий раздел науки.

Пример галлюцинации в ИС 3. Придуман описание несуществующей статьи, которую могли бы написать ученый ag_1 и ученый ag_2 .

Для исключения галлюцинаций возможно накладывать ограничения на возможные отношения между элементами словаря. Проблема цельности (структурированности) онтологии: отсутствие структуры порождает

Проблема размерности онтологии (словаря): слишком большая порождает свиноботанику (свиноводство+ботаника), слишком маленькая отсекает существующие разделы, например, медфизику (медицина+физика).

Проблема большого количества информации: долгие вычисления при большом количестве, глюки при малом количестве

На выходе задача оптимизации для параметров размерности и структурированности онтологии.

Если у ученого стоит задача поиска новых методов для решения определенной задачи или новых моделей для представления задачи, применения моделей к другим предметным областям, поиск данных (описания генерации данных или готовых датасетов) для экспериментов. Это расширение отражения графа онтологии (страница 7 последний абзац)

4.4 Примеры задач анализа информации

Пример 1. Ученый ищет новые методы для решения своей задачи

Пример 2. Ученый ищет данные, на которых можно апробировать алгоритм

Пример 3. Ученый ищет алгоритмы, с которыми можно сравнить его алгоритм

Ситуация 1. Есть группа ученых физиков, которые используют метод X для решения своих задач. Есть группа ученых социологов, которые тоже используют метод X для своих задач. Однако, эти группы не знают о существовании друг друга, поскольку используют разные термины и не находят друг друга, чтобы можно было узнать и сослаться. Это поиск аналогий между разными науками (кстати сюда весело заходит график Новикова о применимости методов в науках, можно на него сослаться). Конкретный пример: управление толпой в социологии и управление атомами идеального газа в физике.

Ситуация 2. Конкуренция и плагиат. Если один ученый украл идею другого и переработал, заменив термины, то как это обнаружить. Допустим, он украл алгоритм, переписал термины и обозначения, укрупнил

некоторые шаги алгоритма или наоборот что-то более подробно раскрыл в шагах. Или украл теорему, немного пересобрал принцип доказательства, переписал иначе результат. Тут также появляется вопрос экстраполяции одной теории в другую.

Ситуация 3. (логическая вероятность). Насколько ошибка или недосказанность определяет результат? Ученый доказал ошибочную теорему, возможно, если там исправить ошибку, то она будет хорошей. Ученый придумал алгоритм, но не доказал его сложность, другой ученый доказал у себя и получил сильный результат. Ученый придумал алгоритм и не знал, что его можно применить в какой-то критической сфере, что это сделает прорыв в науке. Позитивизм и неопозитивизм.

Ситуация 4. Моделирование процесса жизненного цикла научной статьи от написания, рецензирования, внедрения и цитирования, и кто будет цитировать

Далее описание ситуаций, связанных с неполнотой знаний

Ситуация 5. Необходимо обыграть в примерах ситуацию неполного знания. Классический пример « $P = NP$ ». Порождаются многие статьи с доказательством или опровержением данного тезиса, пишутся статьи в ответ на статьи, опровергающие доказательства.

Ситуация 6. Теорема Ферма и ее доказательство. Сформулирована французским математиком Пьером Ферма в 1637 году. Несмотря на простоту формулировки, доказательство теоремы искали многие математики на протяжении более трёхсот лет. Можно в качестве примера описать этапы «приращения знаний» к полному доказательству.

Ситуация 7. Бурбаки: группы и алгебры Ли — новое знание, которое было получено как побочный эффект попытки структурировать всю существующую математику, отказавшись от классических подходов

Список литературы

- [1] Z.K. Avdeeva, M.S. Gavrilov, D.V. Lemtyuzhnikova, and et al. Methods for solving the problem of topic segmentation of texts based on knowledge graphs. *Journal of Computer and Systems Sciences International*, 63:642–662, 2024.
- [2] Akashdeep Bansal, M Balakrishnan, and Volker Sorge. Evaluating cognitive complexity of algebraic equations. *The Journal on Technology and Persons with Disabilities*, 170, 2021.
- [3] Rudolf Carnap. *Logical Foundations of Probability*. Chicago University of Chicago Press, Chicago], 1950.

- [4] Richard T Cox. *Algebra of probable inference*. Johns Hopkins University Press, 2001.
- [5] A. P. Dempster. Technical report no. 20. a generalization of bayesian inference. Contract Nonr 1866(37) NR-042-097, Department of statistics Harvard University a generalization of Bayesian inference, November 15, 1967.
- [6] Didier Dubois, Lluís Godo, and Henri Prade. Weighted logics for artificial intelligence – an introductory discussion. *International Journal of Approximate Reasoning*, 55(9):1819–1829, 2014. Weighted Logics for Artificial Intelligence.
- [7] Jérôme Euzenat, Carlo Allocca, Jérôme David, Mathieu d’Aquin, Chan Le Duc, and Ondrej Sváb-Zamazal. Ontology distances for contextualisation. Contract, INRIA, August 2009. euzenat2009b.
- [8] Alan Fern. Weighted logic. Technical report, College of Engineering, Oregon State University, 2010.
- [9] Nicola Guarino, Daniel Oberle, and Steffen Staab. *What Is an Ontology?*, pages 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [10] Dmitrii Alekseevich Gubanov, Oleg Petrovich Kuznetsov, Evgenii Aleksandrovich Kurako, Darya Vladimirovna Lemtyuzhnikova, Dmitry Alexandrovich Novikov, Aleksandr Gedevanovich Chkhartishvili, et al. Isand: an information system for scientific activity analysis (in the field of control theory and its applications). *Control Sciences*, (3):42–65, 2024.
- [11] Sébastien Harispe, Sylvie Ranwez, Jacky Montmain, et al. *Semantic similarity from natural language and ontology analysis*. Springer Nature, 2022.
- [12] Alan Hájek. Interpretations of Probability. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2023 edition, 2023.
- [13] Inc IPU RAS. ИСАНД — интеллектуальная система анализа научной деятельности. <https://isand.ipu.ru>, 2023. Дата доступа: (05.02.2025).

- [14] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55:1 – 38, 2022.
- [15] Zdzisław Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data. Theory and Decision Library D: (TDLD, volume 9)*. Kluwer Academic, Dordrecht, 1991.
- [16] Rolando Quintero, Miguel Torres-Ruiz, Rolando Menchaca-Mendez, Marco A. Moreno-Armendariz, Giovanni Guzman, and Marco Moreno-Ibarra. Dis-c: conceptual distance in ontologies, a graph-based approach. *Knowl. Inf. Syst.*, 59(1):33–65, April 2019.
- [17] Frank Plumpton Ramsey. *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace and company, New York,, 1931.
- [18] Andrew Schumann and Alexander V. Kuznetsov. Foundations of mathematics under neuroscience conditions of lateral inhibition and lateral activation. *International Journal of Parallel, Emergent and Distributed Systems*, 33(3):237–256, 2018.
- [19] Kristina Šekrst. Unjustified untrue "beliefs": Ai hallucinations and justification logics. In Kordula Świątorzecka, Filip Grgić, and Anna Brozek, editors, *Logic, Knowledge, and Tradition. Essays in Honor of Srećko Kovac*. forthcoming.
- [20] Игорь Александрович Маньковский. Потенциал применения институционализма в качестве теоретической основы государственного управления. *Государственное и муниципальное управление. Ученые записки*, page 59, 2024.
- [21] Агаев Р.П., Алескерев Ф.Т., Алчинов А.И., Барабанов И.Н., Бурков В.Н., Васильев С.Н., Вишневский В.М., Галяев А.А., Губанов Д.А., Калашников А.О., Калянов Г.Н., Каравай М.Ф., Каршаков Е.В., Краснова С.А., Кузнецов О.П., Кульба В.В., Лазарев А.А., Лебедев В.Г., Макаренко А.В., Мандель А.С., Мещеряков Р.В., Михальский А.И., Новиков Д.А., Пашенко А.Ф., Роцин А.А., Толоч А.В., Уткин В.А., Фархадов М.П., Хлебников М.В., Чхартишвили А.Г., and Щепкин А.В. *Теория управления: словарь системы основных понятий*. М.: ЛЕНАНД, 2024.