Донская Анастасия Романовна – e-mail: donsckaia.anastasiya@yandex.ru; кафедра программного обеспечения автоматизированных систем; аспирант.

Ulyev Andrey Dmitrievich – Volgograd State Technical University; e-mail: ulyev-ad@yandex.ru; Volgograd, Russia; the department software engineering; graduate student.

Orlova Yulia Aleksandrovna – e-mail: yulia.orlova@gmail.com; the department software engineering; dr. of eng. sc.; associate professor.

Rozaliev Vladimir Leonidovich – e-mail: vladimir.rozaliev@vstu.ru; the department software engineering; cand. of eng. sc.; associate professor.

Donsckaia Anastasia Romanovna – e-mail: donsckaia.anastasiya@yandex.ru; the department software engineering; graduate student.

УДК 004.8 DOI 10.18522/2311-3103-2023-2-273-280

Х.Б. Штанчаев

НЕСТАТИСТИЧЕСКИЕ МЕТОДЫ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ ИЗ ТЕКСТА

Автоматическое извлечение причинно-следственных связей (ПСС) из текстов естественного языка является сложной проблемой искусственного интеллекта. Большинство первых попыток ее решения подразумевали использование, построенных вручную лингвистических и синтаксических правил на небольших наборах данных. Однако с появлением больших данных, доступной вычислительной мощности и с большим скачком в области машинного обучения, концепция решения данной проблемы постепенно сдвинулась. В данной статье рассмотрена парадигма нестатистического подхода к извлечению причинно-следственных связей, ее основа, языковые конструкции, шаблоны и классификация ПСС. Целью стало исследование методов данной парадигмы определение их недостатков, преимуществ и возможности их применения. В статье рассмотрены различные подходы, приведенные авторами достаточно известных и высоко цитируемых исследовательских работ и их влияние на успешность извлечения причинно-следственных связей. Анализ этих научных работ однозначно подтвердил, что задача извлечения ПСС является крайне сложной задачей обработки естественного языка. Наличие разнообразных лингвистических конструкций языка, двусмысленности различного рода, а также языковые особенности очень сильно влияют на точность извлечения ПСС. Почти все нестатистические методы столкнулись с проблемой узкоспециализированных областей знаний, где почти всегда требуется экспертное описание. Так же практически все нестатистические методы являются ручными или же полуавтоматическими, т. к. предполагают построение шаблонов для определения ПСС в тексте. Несмотря на то, что нестатические методы с достаточной точностью (в среднем 70–80%) успешно справляются с рассматриваемой задачей, на сегодняшний день отсутствует универсальный метод для извлечения ПСС. Предполагаемый метод должен быть универсальным относительно языков, универсальным относительно предметных областей и с возможностью определения неявных ПСС.

Причинно-следственные связи; причинные знания; причинные отношения; обработка естественного языка; двусмысленность; компьютерная лингвистика.

Kh.B. Shtanchaev

NON-STATISTICAL METHODS OF AUTOMATIC EXTRACTION OF CAUSAL RELATIONSHIPS FROM THE TEXT

Most of the first attempts to extraction of causal relationship were tied with complex and manual linguistic patterns, syntactic rules and small datasets based on domain. This article examines the paradigm of a non-statistical approach to the extraction of causal relationships, its basis, language constructs, patterns, and classification of causal relationships. The aim was to study the methods of this paradigm, to determine their disadvantages, advantages, and the possibility of their application.

The article discusses various approaches given by the authors of well-known and highly cited research papers and their impact on the success of the extraction of causal relationships. The analysis of these scientific papers has unequivocally confirmed that the task of extracting CR is an extremely difficult task of natural language processing. The presence of a variety of linguistic constructions of the language, ambiguities of various kinds, as well as language features greatly affect the accuracy of CR extraction. Almost all non-statistical methods have encountered the problem of highly specialized fields of knowledge, where expert description is almost always required. Also, almost all non-statistical methods are manual or semi-automatic, because assume the construction of templates for determining the CR in the text. Even though non-static methods with sufficient accuracy (on average 70-80%) successfully cope with the task under consideration, there is currently no universal method for extracting CR. The proposed method should be universal with respect to languages, universal with respect to subject areas and with the possibility of defining implicit CR.

Causal relationships; causal knowledge; natural language processing; ambiguity; implicit causal relationships; computational linguistics.

Введение. Автоматическое извлечение знаний из текстов на сегодняшний день одна из многих открытых задач в искусственном интеллекте. Эффективно извлечь из текста выражения при постоянно развивающемся словарном запасе языка, не говоря уже о двусмысленных выражениях, метафорах, сарказме и даже неявных отрицаний достаточно непростая задача. Решение данной задачи позволит развить модели искусственного интеллекта до серьезных уровней имитации человеческого мозга. Тем не менее, ученые за последние два десятилетия добились значительных успехов в компьютерной лингвистике. Это позволило создать практически универсальные методы, базирующиеся на извлечении причинно-следственных или временных связей.

В последние годы автоматическое извлечение причинных связей (ПСС) становится все более важной прикладной задачей для таких систем ведения диалога, поиска информации, прогнозирования событий, вывод будущих решений или сценариев и обработки решений. Отношения такие, как например "*часть – целое"*[1], "*если – то"*[2], *"причина – следствие"*[3] позволяют представить важную информацию о том, как различные события и сущности должны восприниматься по отношению друг к другу. В частности, считается, что причинно-следственная связь играет очень важную роль в человеческом познании из-за ее способности влиять на принятие решений.

ПСС исследовались в различных областях наук. Каноническим определением ПСС является логическое выражение вида «*А вызывает Б*» или «*А вызывается тем, что Б*». Многие ученые в своих работах пытались провести классификацию ПСС, для упрощения дальнейших исследований. На сегодняшний день самой популярной является предложенной автором [1].

1. *Причинно-следственные отсылки* используются для связывания определений. Альтенберг классифицировал причинно-следственные отсылки на четыре типа [2]: 1) **наречные ссылки** (наречия), например: *так, следовательно, итак, поэтому* и т.д.[4] 2) **предложные ссылки**, например *из-за*[5] 3) **подчинения** например*: потому что, с, как*[6] и 4) **интегрированные ссылки** такие как: *результатом стал, вот почему, результатом был*[7].

[1] В англ. язычной литературе part – whole.
[2] В англ. язычной литературе if – then.
[3] В англ. язычной литературе cause – effect.
[4] В английском языке so, hence, therefore.
[5] В английском языке because of или on account of.
[6] В английском языке because, as, since.
[7] В английском языке that's why, the result was.

2. *Причинные глаголы* – это переходные глаголы или их формы, которые включают в себя причинный элемент. В качестве примера можно привести в принципе любые глаголы, например *ломать* или *купить*. Их переходные формы заключаются в том, чтобы заставлять *ломаться* и заставлять *покупать*.

3. *Результативные конструкции* – это предложения, в которых для описания состояния объекта используется глагол или глагольная форма. Например, «Я перенес данные с одного носителя на другой».

4. Условные обозначения "*ЕСЛИ - ТО*" часто указывают на то, что предшествующее вызывает последующее.

5. *Причинно-следственные наречия и прилагательные* имеют причинный характер в своих значениях, например, *фатальный* или *катастрофичный*, который может быть перефразирован как *причина смерти*.

Однако следует указать, что классификация подразумевает, что все исследуемые ПСС явные.

Идея нестатистического подхода. Фактически начало нестатистическим подходам положил в 1989 году Селфридж [3]. В своей статье он описывает основные трудности связанные с извлечением ПСС. Автор в своей статье указывает на необходимость понимания методом предметной области для высокой точности извлечения ПСС. Иначе можно упустить значимые ПСС.

В нескольких исследованиях конца 80-х и 90-х годов предпринимались попытки решить проблемы, выделенные Селфриджем. Одна из первых выдающихся работ, в которой была разработана, полнофункциональная автоматическая система выделения причинно-следственных связей из английских энциклопедических текстов была выполнена Капланом и Берри-Рогге в 1991 году [4]. В работе упоминаются фреймы, как основной вид представления знаний. И для работы системы необходимо предварительно подвергать текст семантическому анализу. После анализа система будет выстраивать фреймы ПСС. Сам же модуль извлечения ПСС осуществляет поиск ПСС с помощью 4 методов:

1) использование 20 закодированных вручную подсказок, обозначающих причинно-следственную связь, например «потому что», «из-за», «когда» («because», «due to» и «when»).

2) расширение явных причинно-следственных пар из 1 пункта до причинно-следственных цепочек. Для расширения инициализируется начальное причинно-следственное отношение $X \to Y$. Осуществляется поиск причинной пары $A \to B$ такой, что $B = X$ и причинной пары $\mathrm{C} \to D$ такой, что $Y = C$. Это приводит к цепочке $A \to X \to Y \to D$. Таким методом расширения ПСС можно выстроить цепочки большой длины.

3) Поиск пар ПСС, которые связаны как во времени, так и в пространстве; для поиска авторы используют модель, основанную на работе Дойла [5].

4) ручное построение ограничений для учета специфики предметной области.

У данного подхода имеются следующие ограничения:

- обширная ручная предварительная обработка, такая как кодировка текста по определенным шаблонам и построение подсказок для обнаружения явных и неявных причинно-следственных связей.
- вышеперечисленные шаги возможны если имеются экспертные знания.
- кроме того из пункта 3 неясно каким образом извлекаются понятия, связанные во времени и пространстве. Набор данных невелик и включает не более сотни пояснительных предложений.

Система, которую авторы назвали PROTEUS [6] стала аналогом. Система так же базировалась на делении анализа на две части: *синтаксический* и *семантический*. Обработка текста занимала большое количество времени из-за скудного на-

бора данных. Система так же зависела от предметной области. Авторы Контос и Сидиропулу [7] использовали лингвистические шаблоны для извлечения ПСС. Система работала в ручном режиме. Это предполагало, что шаблоны для текста определялись вручную исходя из предметной области. Работа Гарсии [8] является одной из серьезных работ того времени. Автором была разработана система извлечения ПСС для французского языка, учитывая некоторые его двусмысленности. Несмотря на то, что система осуществляла поиск заранее определенных шаблонов (порядка 23) точность извлечения ПСС достигла 85%. Так же отметим, что система была независимой от предметной области. Вслед за Гарсией большие исследования [1, 9, 10] по извлечению ПСС провел Ху. Почти все работы автора были направлены на выявление только явных ПСС. Автору удалось решить задачу зависимости от предметной области. Тестирование метода проводили в течение четырех месяцев на статьях журнала Wall Street. Общее количество обработанных предложений составило свыше 1100. После обработки результаты сравнивались с двумя экспертами. Система показала 42% точности. Низкая точность связана с невозможностью определения двусмысленностей. Так же шаблоны, которые использовались в работе, были не универсальными и определяли другие конструкции текста не являющимися ПСС.

Ограничения выявленные в процессе работы данного метода, такие как пропуски неявных ПСС, и зависимость от предметной области дорабатывались в другой работе [12]. Метод позволил уйти от большого количества шаблонов необходимых для выявления ПСС. Улучшенная точность и производительность запоминания в этих исследованиях были приписаны алгоритмическим настройкам. Однако это утверждение спорно. Улучшение вполне может быть связано со спецификой предметной области данных. В целом методы извлечения причинно-следственных связей в данные годы можно определить как методы для работы с небольшими текстами зависящих от конкретной предметной области и с ручной подготовкой аннотаций к данным текстам.

Очевидно, что к этой проблеме необходимо было вернуться с более прагматичной точки зрения. Совсем ранние методы рассматривали в качестве решения задачи шаблоны для текста которые были написаны в ручном режиме.

Поэтому последующие исследования были сосредоточены на создании автоматической системы извлечения ПСС с уменьшением влияния предметных областей. Не решенной также оставалась проблема неявных ПСС. Работа [13] опубликованная в 2002 году авторами Гирджу и Молдованом, было следующим исследованием направленное на создание универсального метода извлечения ПСС. Авторами было предложено автоматически выявлять лингвистические шаблоны. Затем осуществлять проверку списка полученных шаблонов на основе семантических ограничений на существительные и глаголы. Чтобы упростить задачу, авторы сосредоточились на явных синтаксических шаблонах вида «NP1-CausativeVerb-NP2» [13][8]. Авторы в своей работе благодаря этому шаблону выделили основные глаголы и провели их ранжирование по двусмысленности и частоте употребления. Для тестирования системы были проведены эксперименты с большим количеством разных новостных статей. Результат работы алгоритма сравнивался с двумя независимыми ручными наборами ПСС. Точность данного алгоритма составила 65,6%. В итоге точность получилась не такая высокая, как например у Ху [9], однако предложенный алгоритм стал первым алгоритмом, который решал, хоть и в полуавтоматическом режиме, проблему двусмысленности.

[8] NP – noun phrase – словосочетание существительного, causative verb – причинный глагол

Попытки решения вышеописанных проблем были предприняты исследователями из Гонконга в серии работ [14–16]. Главная идея этих работ заключалась в создании системы извлечения ППС, основанная на семантических ожиданиях. Авторы предполагали, что люди легко понимают и извлекают ПСС из текстов, потому что у них есть некоторая ожидаемая семантика текста. Она позволяет направлять внимание человека на поиск и понимание информации. Причем ожидаемая семантика отличается исходя из предлагаемого для обработки текста. Например, ожидания от поэзии и научной литературы различные. Система названная SEKE состоит из нескольких типов семантических правил, организованных иерархически. Так как система является полуавтоматической, то на этапе предварительной обработки шаблоны разрабатываются вручную. Согласно этим шаблонам, анализируют весь текст. В работах автора в качестве учебной выборки использовались статьи о фондовом рынке Гонконга в количестве более 360 штук. Следует указать, что шаблоны для извлечения разрабатываются от простых к сложным предложениям. Система обрабатывает текст, отсеивая все предложения, не связанные с причинностью. Следующим шагом идет обработка оставшихся предложений на причины и следствия путем сопоставления с построенными шаблонами. Система проверяет предложения еще раз на предмет вхождения простых ПСС в более сложные. Если какая-то фраза соответствует нескольким шаблонам, то ее ранжируют по соответствующим критериям. Функция подсчета ранга тщательно подобрана авторами с учетом поддержки шаблонов. Однако производится она вручную. Так же для исключения двусмысленности система SEKE включает компонент на основе WordNet. Он используется для поиска понятий, которые являются синонимами к извлеченным ПСС и фразам. Так же такой подход позволяет сгенерировать новые ПСС, которые ранее были неявными. Проработав статьи, связанные с фондовым рынком авторы добились точности работы системы более 70%. В процессе тестирования авторы обнаружили, что их система определила некоторые неявные шаблоны. Недостатком системы явилось то, что авторы не отвязали систему от одной конкретной предметной области, поэтому неясно, как система будет масштабироваться для различных текстов.

Рассмотренные выше методы установили некий компромисс между высокой точностью и возможностью применения систем для нескольких предметных областей. Разработанные меры для семантического ранжирования оказались наиболее популярным методом обработки неоднозначных причинно-следственных связей. Преимущественно, никто не уделял особого внимания выделению неявных причинно-следственных связей, которые не имеют конкретной лингвистической формы, но все же влияют на читателя, чтобы причинно связать два или более события в текст.

Итту и Боума в своих работах [17] и [18] попытались решить эту сложную проблему, вернувшись к текстам, специфичным для предметной области, в качестве меры упрощения. Авторы разделили и сделали упор на трех типах неявных ПСС:

1) результирующий «вербальный» шаблон, определяющий конкретную ситуацию (например глаголы – уменьшить, стрелять);

2) шаблон, который делает причинные факторы неотделимыми от ситуации (например «яркий свет засвечивает пленку»);

3) «невербальный» шаблон (например, из-за).

Кроме того, они предоставили способ устранить неоднозначность многозначных шаблонов. Такие шаблоны имеют множество значений и затрудняют извлечение ПСС. Например, глагол «ведет» в следующем предложении имеет причинный смысл «курение приводит к раку легких», а в этом «тропинка ведет в сад», нет. Конкретно, их методика использует Википедию для создания базы знаний,

поскольку она имеет широкий охват и с высокой вероятностью содержит широкий спектр явных и неявных лингвистических шаблонов, кодирующих причинно-следственную связь. Сначала извлекаются все самые короткие шаблоны, связывающие две словосочетания с существительными. Чтобы извлечь причинно-следственные связи из этого списка общих связей, предлагаемый алгоритм начинается с начальной причинно-следственной пары (например, в работе авторов используется вич - спид) и идентифицирует шаблоны в Википедии, которые соединяют начальную пару. Вычисляется показатель надежности для каждого шаблона, компилируются *k* самых надежных шаблонов, а затем рекурсивно идентифицируются другие пары, связанные этими шаблонами. Количество итераций данной процедуры вызываются для всех пар шаблонов. Функция же выполняется до тех пор, пока не будут извлечены все определенные ПСС. Алгоритм проверялся авторами на так называемой *выгрузке* статей из Википедии (порядка 500 млн фраз). Конечная точность была определена равной 76%. К преимуществам метода можно отнести извлечение не явных ПСС после подстройки шаблонов и выполнения большого количества итераций на *k* самых надежных шаблонах.

Более поздние работы основанные на нестатистических методах, были посвящены исследованиям и разработке более сложных шаблонов, которые являются как временными, так и причинно-следственными [19, 20]. Были попытки автоматического определения ПСС в арабском языке [21]. В некоторых статьях, посвященных конкретным приложениям, рассматривается извлечение причинно-следственных связей из новостных тем [22–24], и нахождение причинно-следственной связи в специализированной области окружающей среды [25]. Но данные методы так же не имеют независимости от предметной области, используются для извлечения ПСС в каком-то конкретном языке и основаны на шаблонных методах.

Выводы. В статье представлен обзор исследовательской литературы, по ранним нестатистическим методам автоматического извлечения причинно-следственных. Кроме того, в статье приведено подробное обсуждение по крайней мере пяти высоко цитируемых исследований. Анализ данных работ позволил однозначно определить, что извлечение причинно-следственных связей является одной из самых сложных задач обработки естественного языка, прежде всего, из-за наличия лингвистических конструкций, которые могут быть причинными, а могут и не быть, в зависимости от текстового контекста. На извлечение ПСС сильно влияют так же двусмысленности и наличие неявных причинно-следственных связей. Еще одна проблема — это узкоспециализированные области знаний, где почти всегда требуется экспертная аннотация и описание вручную неких шаблонов, что затрудняет разработку универсальных методов решения этой проблемы. Стоит отметить, что хорошие результаты точности, достигнутые учеными в последнее время, обусловлены тем, что до начала работы почти всех вышеописанных алгоритмов вручную строятся шаблоны. Также почти все методы работают конкретно для английского языка с его лексической базой.

Учитывая описанное, можно сделать следующие однозначные выводы:

1) нестатистические методы могут успешно использоваться для извлечения ПСС из каких-то конкретных текстов, связанных с одной конкретной предметной областью, заранее проработав вручную шаблоны для извлечения;

2) на сегодняшний день отсутствует метод для извлечения ПСС, универсальный относительно языков, универсальный относительно предметных областей и с возможностью определения неявных ПСС;

3) с большой долей вероятности методы данного подхода проиграют методам, основанным на машинном обучении, нейронных сетях и нечеткой логике.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Khoo C.S., Kornfilt J., Oddy R.N., Myaeng S.H.* Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing, *Literary and Linguistic Computing*, 1998. Vol. 13, No. 4, pp. 177-186.
2. *Altenberg B.* Causal linking in spoken and written English, *Studia linguistic*, 1984. Vol. 38, No. 1, pp. 20-69.
3. Selfridge M. Toward a natural language-based causal model acquisition system // Applied Artificial Intelligence an International Journal. 1989. Vol. 3, №2-3. P. 191–212.
4. *Kaplan R. M., Berry-Rogghe G.* Knowledge-based acquisition of causal relationships in text // Knowledge Acquisition, 1991, Vol. 3, No. 3, pp. 317-337.
5. *Doyle R.J.* Hypothesizing and refining causal models, *DTIC Document, Tech. Rep*., 1984.
6. *Grishman R.* Domain modeling for language analysis, *Linguistic approaches to artificial intelligence*, 1990, pp. 41-58.
7. *Kontos J., Sidiropoulou M.* On the acquisition of causal knowledge from scientific texts with attribute grammars, *International Journal of Applied Expert Systems*, 1991, Vol. 4, No. 1, pp. 31-48.
8. *Garcia D.* Coatis, an NLP system to locate expressions of actions connected by causality links, in Knowledge Acquisition, Modeling and Management, Springer, 1997, pp. 347-352.
9. *Khoo C.S., Chan S., Niu Y.* Extracting causal knowledge from a medical database using graphical patterns, *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics*, 2000, pp. 336-343.
10. *Khoo C.S., Chan S., Niu Y., Ang A.* A method for extracting causal knowledge from textual databases, *Singapore journal of library & information management*, 1999, Vol. 28, pp. 48-63.
11. *Khoo C.S.* Automatic identification of causal relations in text and their use for improving precision in information retrieval // Ph.D. dissertation. 1995.
12. *Finn R.* Program uncovers hidden connections in the literature, *The Scientist*, 1998, Vol. 12, No. 10, pp. 12-13.
13. *Girju R. and Moldovan D.* Text mining for causal relations, *FLAIRS Conference*, 2002, pp. 360-364.
14. *Chan K. and Lam W.* Extracting causation knowledge from natural language texts, *International Journal of Intelligent Systems*, 2005, Vol. 20, No. 3, pp. 327-358.
15. *Low B.-T., Chan K., Choi L.-L., Chin M.-Y., Lay S.-L.* Semantic expectation-based causation knowledge extraction: A study on Hongkong stock movement analysis, *Advances in Knowledge Discovery and Data Mining*. Springer, 2001, pp. 114-123.
16. *Chan K., Low B.-T., Lam W., Lam K.-P.* Extracting causation knowledge from natural language texts. Springer, 2002.
17. *Ittoo A., Bouma G.* Extracting explicit and implicit causal relations from sparse, domain-specific texts, *Natural Language Processing and Information Systems*. Springer, 2011, pp. 52-63.
18. *Ittoo A., Bouma G*. Minimally supervised learning of domain specific causal relations using an open-domain corpus as knowledge base, *Data & Knowledge Engineering*, 2013, Vol. 88, pp. 142-163.
19. *Bethard S., Corvey W.J., Klingenstein S., Martin J.H.* Building a corpus of temporal-causal structure, *LREC*, 2008.
20. *Al Hashimy H., Saleh A., Kulathuramaiyer N.* Ontology enrichment with causation relations, *Systems, Process & Control (ICSPC) 2013 IEEE Conference on*. IEEE, 2013, pp. 186-192.
21. *Sadek J.* Automatic detection of arabic causal relations, *Natural Language Processing and Information Systems*. Springer, 2013, pp. 400-403.
22. *Ackerman E.J.M.* Extracting a causal network of news topics, *On the Move to Meaningful Internet Systems: OTM 2012 Workshops*. Springer, 2012, pp. 33-42.
23. *Ishii H., Ma Q., Yoshikawa M.* Causal network construction to support understanding of news, *in System Sciences (HICSS), 2010 43rd Hawaii International Conference on*. IEEE, 2010, pp 1-10.
24. Incremental construction of causal network from news articles Information and Media Technologies, 2012, Vol. 7, No. 1, pp. 110-118.
25. *Arauz P.L. and Faber P.* Causality in the specialized domain of the environment, *Semantic Relations-II. Enhancing Resources and Applications Workshop Programme*, Citeseer, 2012, pp. 10.

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Целых

Штанчаев Хайрутин Баширович – Дагестанский государственный технический университет; e-mail: shtanchaev.h@gmail.com; г. Махачкала, Россия; тел.: +79883081572; кафедра ПОВТиАС; к.т.н.

Shtanchaev Khairutin Bashirovich – Dagestan State Technical University; e-mail: shtanchaev.h@gmail.com; Makhachkala, Russia; phone: +79883081572; the department of POVTiAS; cand. of eng. sc.

УДК 004.832.23 DOI 10.18522/2311-3103-2023-2-280-298

С.И. Родзин

СОВРЕМЕННОЕ СОСТОЯНИЕ БИОЭВРИСТИК: КЛАССИФИКАЦИЯ, БЕНЧМАРКИНГ, ОБЛАСТИ ПРИМЕНЕНИЯ*

Целями данной статьи является анализ современного состояния исследований в области разработки алгоритмов, инспирированных природой, включая категоризацию, классификацию, тестирование, цитируемость и области применения. Представлена новая многоуровневая система классификации на основе следующих признаков: критерий соответствие природной метафоре, структурный, поведенческий, поисковый, компонентный и оценочный критерии. Классификация биоэвристик предполагает систематическое отнесение каждой биоэвристики к одному и только одному классу в рамках системы взаимоисключающих и неперекрывающихся классов. Категоризация позволяет объективно подходить к выбору биоэвристик. Для каждой биоэвристики имеются конкретные задачи, с которыми она хорошо справляется. Знать эти взаимосвязи важно для целенаправленного применения биоэвристики. Рассмотрен пример классификации. Отмечено, что наиболее информативным критерием классификации является поведенческий критерий, наиболее цитируемым классом биоэвристик являются алгоритмы роевого интеллекта, а наиболее цитируемой биоэвристикой – алгоритм роя частиц PSO. Представлены современные подходы к бенчмаркингу биоэвристик: задачи дискретной и непрерывной оптимизации, а также оптимизационные инженерные задачи. Отмечена тенденция проводить сравнение производительности биоэвристик, используя статистическую проверку гипотез на бенчмарках. Систематизированы задачи, успешно решаемые биоэвристиками в таких областях, как инженерное проектирование, обработка изображений и компьютерное зрение, компьютерные сети и коммуникации, энергетика и энергоменеджмент, анализ данных и машинное обучение, робототехника, медицинская диагностика. Наметилась тенденция к гибридизации биоэвристик в одном оптимизаторе. Однако требуются убедительные доказательства, что результаты компенсируют увеличение сложности по сравнению с отдельными алгоритмами. Отмечены задачи оптимизации, требующие дальнейших исследований: задачи динамической и стохастической оптимизации; задачи многокритериальной оптимизации; задачи мультимодальной оптимизации; задачи многомерной оптимизации; задачи меметической оптимизации, в которых комбинируется множество поисковых алгоритмов; задачи оптимизации и адаптации настроек параметров биоэвристик для достижения баланса между скоростью сходимости и диверсификацией пространства поиска решений.

Биоэвристика; классификация; категоризация; бенчмаркинг; фреймворк; агент; оператор; популяция; стигмергия.

* Исследование выполнено за счет гранта Российского научного фонда № 23-21-00089, https://rscf.ru/project/23-21-00089/ в Южном федеральном университете.