Математические структуры и моделирование 2021. № 1(57). С. 108–113

УДК 621.396+004.8
DOI 10.24147/2222-8772.2021.1.108-113

СТЕГАНОГРАФИЯ МАТЕМАТИЧЕСКИХ ТЕКСТОВ С ОШИБКАМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Д.Э. Вильховский
ассистент, e-mail: vilkhovskiy@gmail.com
А.К. Гуц
д.ф.-м.н., профессор, e-mail: guts@omsu.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. Рассматривается использование в стеганографии в качестве скрытого сообщения ошибок в математических формулах, примерах решения задач или в доказательствах теорем. Обсуждается применение систем искусственного интеллекта для извлечения скрытой информации из стегоконтейнера.

Ключевые слова: стеганография, математические тексты с ошибками, искусственный интеллект.

В заметке [1] предлагалось использовать математические тексты для скрытой передачи битов информации в форме математических формул. Ошибка в формуле – это бит 0, формула без ошибки – 1. Весь текст, таким образом, несёт скрытую последовательность битов: 011100101....

С точки зрения стеганографии мы скрываем некоторый код, состоящий из последовательности нулей и единиц, в математическом тексте, который является *контейнером*, а этот же математический текст, но с ошибками — *стегоконтейнером*.

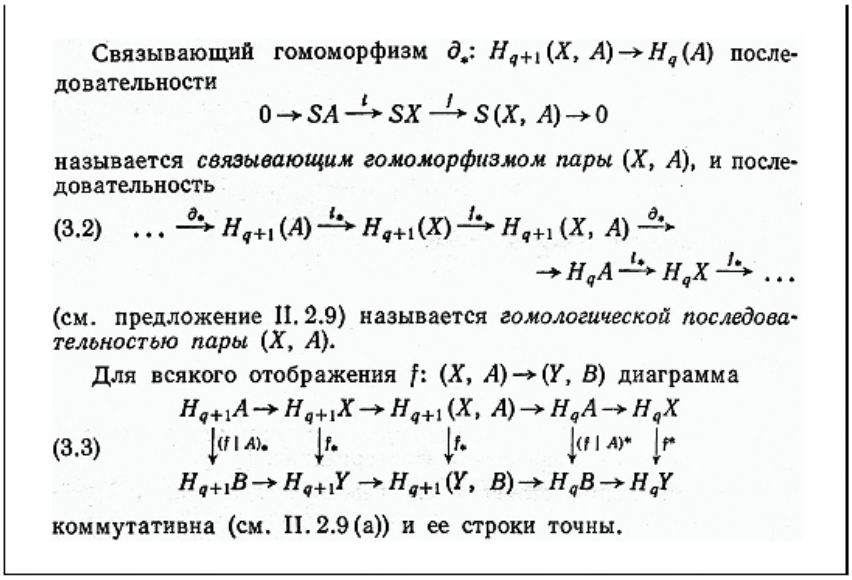
Очевидно, что любой математический текст может быть прочитан и понят только специалистом в конкретной области математики. Другими словами, для извлечения скрытого сообщения из стегоконтейнера требуется специалист — человек со специальным образованием и достаточно высокой квалификацией.

Действительно, ошибки в формулах (в геометрических рисунках) в тексте, относящемся к элементарной математике, легко находятся выпускником средней школы, имевшим отличные оценки по математическим дисциплинам. Однако проверка текста по римановой геометрии требует уже отменных знаний в данной науке. И если знатока римановой геометрии можно ещё найти в вузах областного центра, то проверка корректности текста в области алгебраической геометрии, или гомотопической топологии, или по теории топосов потребует, во-первых, осознания самого факта, что он относится к данным наукам, а, во-вторых, будет связана с весьма хлопотными поисками человека, способного его проверить, поскольку не в каждом регионе (стране!), такой нужный специалист найдётся.

1. Примеры сокрытия сообщения

Ошибка может быть в формуле, в формулировке теоремы, в примере, в иллюстрации и т. д. Места сокрытия сообщения — это, к примеру, договорённость допускать ошибки только в пронумерованных формулах. Очевидно, знание такого ключа резко сокращает время извлечения информации из полученного текста. Ошибка в пронумерованной формуле — это бит 0, следующая пронумерованная формула без ошибки — 1.

Весь текст, таким образом, несёт скрытую последовательность битов: 011100101...



Связывающий гомоморфизм ∂_*: $H_{q+1}(X, A) \to H_q(A)$ последовательности

$$0 \to SA \xrightarrow{i} SX \xrightarrow{j} S(X, A) \to 0$$

называется *связывающим гомоморфизмом пары* (X, A), и последовательность

$$(3.2) \quad \ldots \xrightarrow{\partial_*} H_{q+1}(A) \xrightarrow{i_*} H_{q+1}(X) \xrightarrow{j_*} H_{q+1}(X, A) \xrightarrow{\partial_*} \to H_qA \xrightarrow{i_*} H_qX \xrightarrow{j_*} \ldots$$

(см. предложение II. 2.9) называется *гомологической последовательностью пары* (X, A).

Для всякого отображения f: $(X, A) \to (Y, B)$ диаграмма

$$(3.3) \quad \begin{array}{ccccccccc} H_{q+1}A & \to & H_{q+1}X & \to & H_{q+1}(X, A) & \to & H_qA & \to & H_qX \\ \downarrow{\scriptstyle (f\,|\,A)_*} & & \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle f_*} & & \downarrow{\scriptstyle (f\,|\,A)^*} & & \downarrow{\scriptstyle f^*} \\ H_{q+1}B & \to & H_{q+1}Y & \to & H_{q+1}(Y, B) & \to & H_qB & \to & H_qY \end{array}$$

коммутативна (см. II. 2.9 (а)) и ее строки точны.

Рис. 1. Пример текста из алгебраической топологии

На рис.1 приводится пример простейшего текста из теории гомологий (раздел алгебраической топологии). В него легко внести ошибки: достаточно повернуть в обратном направлении вертикальные стрелки в формуле (3.3) или вместо f^* написать f_*. Этот пример показывает, что даже если воспользоваться возможностями современного интернета и обратиться к специалисту-топологу в другой стране, то ему потребуется значительное время на проверку текста. Замечательная особенность приведённого примера, взятого из учебника А. Дольда «Лекции по алгебраической топологии», заключается в том, что в формуле (3.3) имеются две опечатки: самые правые символы * должны стоять не как верхние индексы, а как нижние. Следовательно, нельзя доверить проверку правильности формулы (3.3) *человеку*, незнакомому с теорией гомологий. Таким образом, поскольку в нумерованной формуле (3.3) имеются две ошибки, а формула (3.2) не содержит ошибок, то в свете данного тезиса можно заявить, что формулы (3.2), (3.3) несут три бита информации — 100, а учебник А. Дойча, безобидно стоящий на книжной полке, хранит в себе ключ для выявления скрытых сообщений.

В текстовой стеганографии существуют специальные программы, способные вносить скрытную информацию в текстовые файлы, например в виде определённого количества пробелов в конце каждой строки. Нетрудно создать программу, порождающую тексты по математике (физике, химии), способные тайно прятать информацию в форме ошибок в формулах.

Стегоконтейнеры, построенные на принципе включения ошибок в формулах, назовём *интеллектуальными*.

Недостатки интеллектуальных стегоконтейнеров:

1) малый объём передаваемого файла, позволяющий злоумышленнику быстрее его обрабатывать; однако можно искусственно увеличивать объём, добавляя к стегоконтейнеру несколько книг по используемой тематике;

2) описки автора в формулах скрытого сообщения; опечатки в формулах, допущенные типографией и пр., которые засоряют скрытое сообщение и вносят искажения при его извлечении.

2. Интеллектуальные системы извлечения сообщений

Как уже говорилось, извлекать сообщение из интеллектуального стегоконтейнера сможет только специалист высочайшего уровня. Можно ли привлечь для этого машины, точнее, компьютеры или, если заглядывать вперёд, системы с искусственным интеллектом?

Очевидно, что можно создать *базу данных*, содержащую всю литературу по алгебраической топологии и перебирать её с помощью суперЭВМ в поиске близкого к интеллектуальному стегоконтейнеру текста. Можно путём сравнения выявлять ошибки и тем самым выявлять скрытые сообщения.

Однако этого мало. Достаточно часто математики используют собственные обозначения вместо распространённых. Имеются также обозначения, принятые в конкретной научной школе и противоположные тем, что используются повсеместно в данное время[1]. Кроме того, при закладке сообщения могут применяться приёмы, затрудняющие поиск похожих текстов в базе данных.

Но можно сообщение скрывать не только в форме ошибки в формуле, но и в форме неверных теорем, либо в форме тупиковых подсказок к доказательствам верных теорем, или ошибок в решениях задач, или в тексте доказательства теорем намеренно делаются ошибочные выводы.

Следовательно, база данных должна включать различные комментарии к текстам и всевозможные исторические справки, касающиеся научных школ и авторов книг. Иначе база данных должна давать *знания*.

Знания — это не только информация об объектах и ситуациях окружающей среды, но и рецепты того, как поступать, т. е. совершать действия при появлении в поле деятельности объектов, и при иных ситуациях. Знания не только

[1]Например, в учебниках Л. Дандау и Е. Лифшица «Теория поля» и П. Рашевского «Риманова геометрия и тензорный анализ» — очень популярных в СССР — латинские индексы $i, k, \ldots$ при изложении теории гравитации Эйнштейна принимают значения 0,1,2,3, а греческие индексы 1,2,3. В западной литературе строго наоборот.

накапливаются, т. е. поступают из среды и запоминаются, но и возникают при обучении людей, а для нас важно добавить — и при *обучении машины* — совершать действия посредством приобретения опыта и адаптации к меняющимся обстоятельствам.

Мы приходим к выводу: чтобы разобраться с самыми сложными скрытыми сообщениями, нужна уже не просто база данных, а *база знаний*.

Почему? Уточним отличия баз знаний от баз данных [2]:

• *Знания более структурированы*, т. е. имеются различные элементы, распределённые по уровням иерархии и имеющие определённые содержания и цели; есть связи и отношения между элементами, поддерживающие те или иные концепции и методы, эти связи и отношения образующие; присутствуют алгоритмы, обеспечивающие существование и надёжность базы знаний.

• *В базе знаний наибольшее значение имеют не атомарные элементы (единицы) знаний (как в базе данных), а взаимосвязь между ними*.

• *Знания более самоинтерпретируемы, чем данные*, т. е. в знаниях содержится информация о том, как их использовать.

• *Знания активны* в отличие от пассивных данных, т. е. знания могут порождать действия системы, использующей их.

Как видим, базы знания — это то, что должно способствовать поиску ошибок в математических текстах. На базах знаний основаны прикладные системы, реализующие логическое направление в построении искусственного интеллекта. Но поскольку речь идёт и об обучении машины, то необходимо задействовать и другое направление в построении *искусственного интеллекта*, помимо направления *представления знаний*, а именно — обучаемые *нейронные сети*.

Искусственный интеллект (artificial intelligence) — условное обозначение кибернетических систем, моделирующих некоторые стороны интеллектуальной[2] деятельности человека — логическое, аналитическое мышление [3].

В технической кибернетике под созданием искусственного интеллекта понимается моделирование интеллектуальной деятельности в искусственных средах с помощью вычислительных машин. Иначе говоря, строится машина, которая в определённых сферах человеческой деятельности, для которых машины сделана и где требуется принятие правильных решений, способна заменить человека. Например, либо полностью заменить пилота авиалайнера или шофёра автобуса (такси), либо обучать детей математике и т. д. В нашем случае допускается, что машина может заменить очень талантливого математика, специалиста в крайне абстрактных и сложных областях математики.

Машина, наделённая искусственным интеллектом, должна обладать такими способностями мозга, как решение задач путём приобретения, запоминания и целенаправленного преобразования *знаний* в процессе обучения на опыте и адаптации к разнообразным обстоятельствам, возникающим в среде, окружающей машину [3].

---

[2]*Интеллект* (лат. intellectus) — означает ум, рассудок, разум, мыслительные способности человека.

[3]Лопатников Л.И. Экономико-математический словарь. – М.: Дело, 2003.

Но всему ли можно научить машину?

3. Обучаемость может быть неразрешимой

Так называется статья израильских математиков, в которой показывается, что обучаемость подобна доказуемости: как не всё можно доказать, так и не всему можно в математике научить машину. Авторы описывают простые сценарии, в которых обучаемость не может быть доказана или опровергнута с помощью стандартных аксиом математики [4]. Примером оказалась задача «оценка максимума».

Хотя, быть может, это касается принципа современных систем машинного обучения и искусственного интеллекта. Они постепенно учатся «видеть» определённые закономерности и отличать правильные ответы от неправильных, используя обширные базы знаний, подготовленные человеком [5].

Если же метод обучения не принципиален, то предложенный в статье способ скрытия информации окажется стеганостойким.

Литература

1. Гуц А.К., Вахний Т.В. Применение математических текстов с ошибками в стеганографии // Межвузовская научно-практическая конференция «Информационные технологии и автоматизация управления». Омск : ОмГТУ, 2009. С. 168–169.
2. Системы искусственного интеллекта / сост. А.В. Гаврилов. Новосибирск : НГТУ, 2004. 74 с.
3. Гуц А.К. Кибернетика. Омск : Изд-во ОмГУ, 2014. 188 с.
4. Ben-David S., Hrubes P., Moran S., Shpilka A., Yehudayoff A. Learnability can be undecidable // Nature Machine Intelligence. 2019. V. 1. P. 44–48.
5. Математики усомнились во всемогуществе искусственного интеллекта [Электронный ресурс]. URL: `https://ria.ru/20190109/1549132155.html` (дата обращения: 20.01.2021)

STEGANOGRAPHY OF ERRONEOUS MATHEMATICAL TEXTS AND ARTIFICIAL INTELLIGENCE

D.E. Vilhovsky
instructor, e-mail: vilkhovskiy@gmail.com
A.K. Guts
Dr.Sc. (Phys.-Math.), Professor, e-mail: guts@omsu.ru

Dostoevsky Omsk State University

Abstract. The use in steganography of errors in mathematical formulas, in examples of problem solving or in theorem proving as hidden messages is considered. Application of artificial intelligence systems to extract hidden information from stegocontainer is discussed.

Keywords: steganography, mathematical texts with errors, Artificial Intelligence.

References

1. Guts A.K. and Vakhnii T.V. Primenenie matematicheskikh tekstov s oshibkami v steganografii // Mezhvuzovskaya nauchno-prakticheskaya konferentsiya ¡¡Informatsionnye tekhnologii i avtomatizatsiya upravleniya¿¿, Omsk, OmGTU Publ., 2009, pp. 168–169. (in Russian)
2. Sistemy iskusstvennogo intellekta. sost. A.V. Gavrilov. Novosibirsk, NGTU Publ., 2004, 74 p. (in Russian)
3. Guts A.K. Kibernetika. Omsk, Izd-vo OmGU, 2014, 188 p. (in Russian)
4. Ben-David S., Hrubes P., Moran S., Shpilka A., and Yehudayoff A. Learnability can be undecidable. Nature Machine Intelligence, 2019, vol. 1, pp. 44–48.
5. Matematiki usomnilis' vo vsemogushchestve iskusstvennogo intellekta [Elektronnyi resurs]. URL: `https://ria.ru/20190109/1549132155.html` (20.01.2021) (in Russian)

Дата поступления в редакцию: 22.01.2021