

International Knowledge Graph Reasoning Challenge (IKGRC)

Application Sheet

1.Information about the applicants

- Authors [Name or Team's name]: Shotaro hattori , Akihiro Fujii
- Affiliations: Hosei University
- E-mail (contact): shotaro.hattori.6z@stu.hosei.ac.jp
fujii@hosei.ac.jp

2. Explanation of the reasoning and estimation process

Methods of Assuming the Culpability of the Culprit

2.1 Preprocessing

First, we downloaded a total of 20 mystery novels (Sherlock Holmes series) from Aozora Bunko. Next, for each of the 20 novels, we programmatically removed all noise, such as ruby in the text and the numbers of the illustrations.

Figure 1 shows the 20 downloaded works.

From these 20 pre-processed works, two works are actually selected for similarity comparison.

番号	小説名	番号	小説名	番号	小説名	番号	小説名
0	ボヘミアの醜聞	5	サセックスの吸血鬼	10	白銀の失踪	15	最後の事件
1	まだらのひも	6	赤毛連盟	11	同一事件	16	三枚の学生
2	グロリアスコット号	7	自転車乗りの影	12	ライギットパズル	17	背中の曲がった男
3	悪者同屋人	8	黄色の顔	13	雲のなじめた男	18	蒼炎石
4	株式仲間人	9	踊る人形	14	空家の冒険	19	槍の種五粒

Figure 1 List of downloaded works(Japanese)

2.2. Scene Segmentation

The number of characters in the main text of each novel was counted and divided into 10 scenes so that the number of characters would be divided into 10 sections. In cases where the text did not follow the intent or where the number of characters could not be neatly divided into ten, adjustments were made as necessary.

Figure 2 below shows the result of dividing "A String of String" into 10 scenes.

ふと私の覚え書きを見ると、七〇もの不思議な事件がある。この八年間、事件を通じて友人シャーロック・ホームズ(とほど、深い楽しみを感じるものはなかった。あの迅速な推理に驚くのもよい。早速きて当てずっぽうにも思えるが、てくれる者もありますが、あいにく、力にならないのです。

そこでホームズさんのことを思い出して、ファリントッシュさんからうかがって、助けていただいたことがあると。つまりこれが義理の父なのですが、そこから抜け出ようと思っただけで、頼みから学資を立て替えてもらい、どうにしたら恐ろしく力の強い人で、怒ったとなると全く手がつけられなくなるのでございます。

つい先週も、父は村の鍛冶屋さんを橋の欄干から川の中へ突き落としてしまいました。訴え出るというのを、かき集めております。寮室同士は行き来ができませんが、各部屋の入り口は同じ廊下に並んでおります。こんな説明でよろしいかと、姉の言ったとおりの低い口笛が一声、聞こえたような気がしました。

それに続いて、がちゃんと何か重い金物でも落ちたような音がしました。廊下を駆け出していきますと、姉の部屋で、夜にはいつも太い鉄の棒をはめて閉ざしております。四方の壁も全部叩いてみましたが、どこもしっかり堅い感じがします。」「賢明な行動です。お話はそれですべて？」

「ええ。」「お嬢さん、まだあります。父親のことをお隠しです。」「どういことでしょうか？」返事の代わりに、るというはっきりした事実。死に際の『ひも』という言葉の謎。それから最期にヘレン・ストーナの聞いた金属音(。老医師は吠えたあと、曲がった火掻き棒を暖炉の中へ放り込み、大手を振って部屋から出ていった。

「ずいぶん愛嬌のある人物だ。」と、ホームズは笑い出しながら、「僕も身体は大きくないが、待っていれば決して行広がり、頂上は深い緑に覆われた森になっていた。その木々の間から、古屋敷の灰色の破風と高い屋根がのぞいて見えたとき、悪人の姿はどこにも見えなかった。

ホームズは手入れの行き届いていない芝地をゆっくりと歩き回り、窓を外側からつぶさに調べた。「見たところ、この姉の部屋へ通じております。」「他のものより真新しいようですが。」「ええ、二年前につけたばかりです。」「ずいぶることが一つ増えました。」それからホームズは木の椅子の前にしゃがんで、背の部分を念入りに調べた。

「感謝します。はっきりしてきました。」と、ホームズは立ち上がり、拡大鏡を懐にしまいながら、「おや、これは何なのかどうか。姉は恐怖のあまり亡くなったのですか？」「いや、違うでしょう。より判然とした原因があるかと。台どこにある？」

「どこにもない。」「ご婦人は寝台を動かせなかった。いつも同じ位置関係だったんだ、通風口と網に対して――そう。通風口から漏れて、さとられるおそれが。」もう一度うなずく。「眠るのもしない。君の命に関わる。いつでもいも消え入り、あたりは元の静けさに戻った。

「何だ、今のは。」私は息を呑んだ。「すべてが終わった知らせだ。」とホームズが答える。「そう、この結末が最も一例だ。ロマがいた。『ひも』という言葉の謎を哀れなご婦人が口走った。つまりそれはマッチを擦ったとき、一瞬何が起けると言えるが、さりとて、大した良心の呵責を感じない。

Figure 2: Divided into 10 scenes(Japanese)

2.3. Similarity Calculation

In order to extract the scenes in which the culprit appears using BERTscore, the following two conditions were paid attention to and used to calculate the similarity.

- (1). The scene in which the culprit is seen for the first time.
- (2). (2). The scene where the crime was committed.

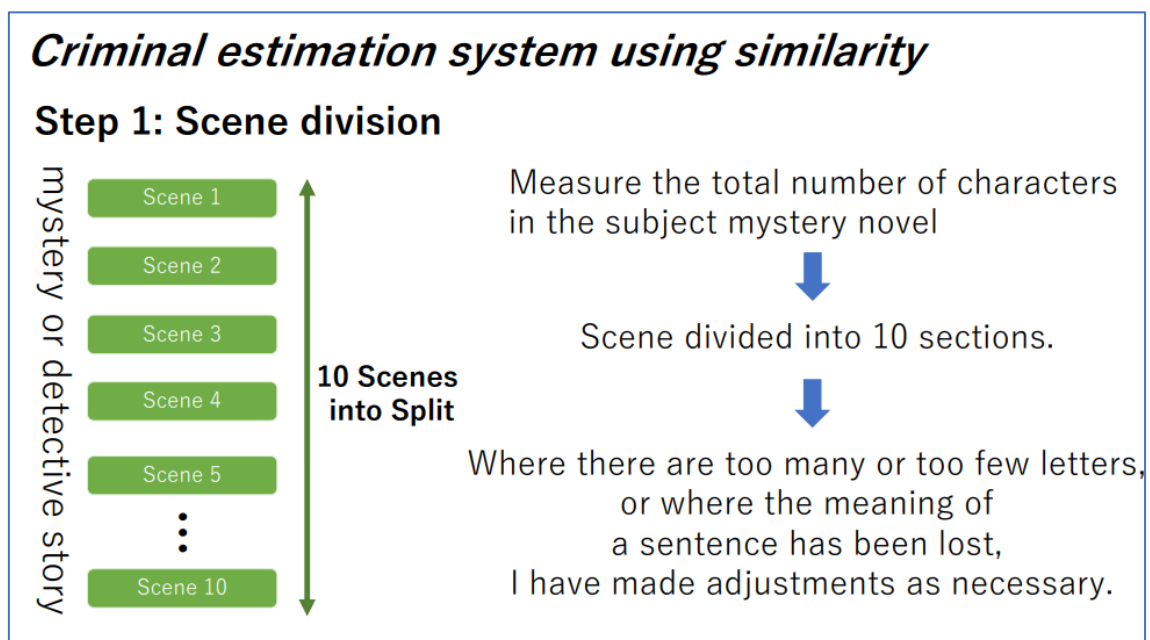
The above conditions (1) and (2) are information that must be obtained in advance from one of the two novels selected from the 20 novels in 4-1.

2.4. Extraction of unique expressions

We used spaCy to extract the names of the people in the scenes of other works that have the highest similarity to the scenes we focused on in 2.3. If the extracted person's name is the culprit, the goal is achieved. The following four conditions are used in the name extraction process.

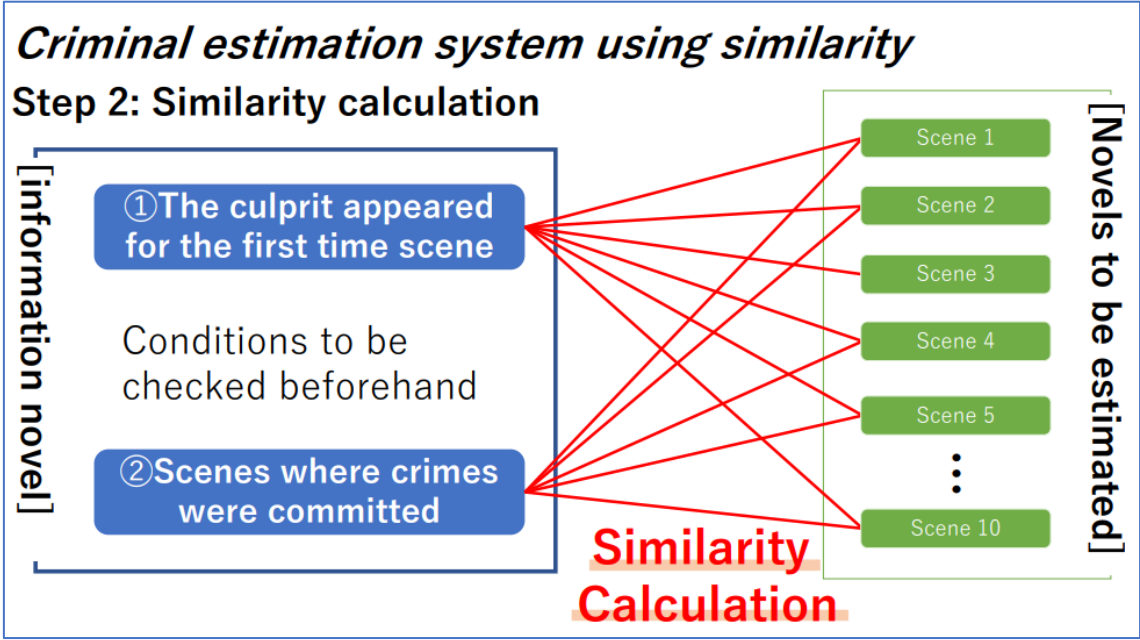
1. Only nouns labeled "PERSON" are displayed among the extracted nouns.
2. exclude major characters who cannot be the killer, such as Holmes and Watson
3. added a process to exclude Chinese characters, which are considered to be noise
4. spacy's Japanese model was used as the language model

2.2 Scene Division Illustration



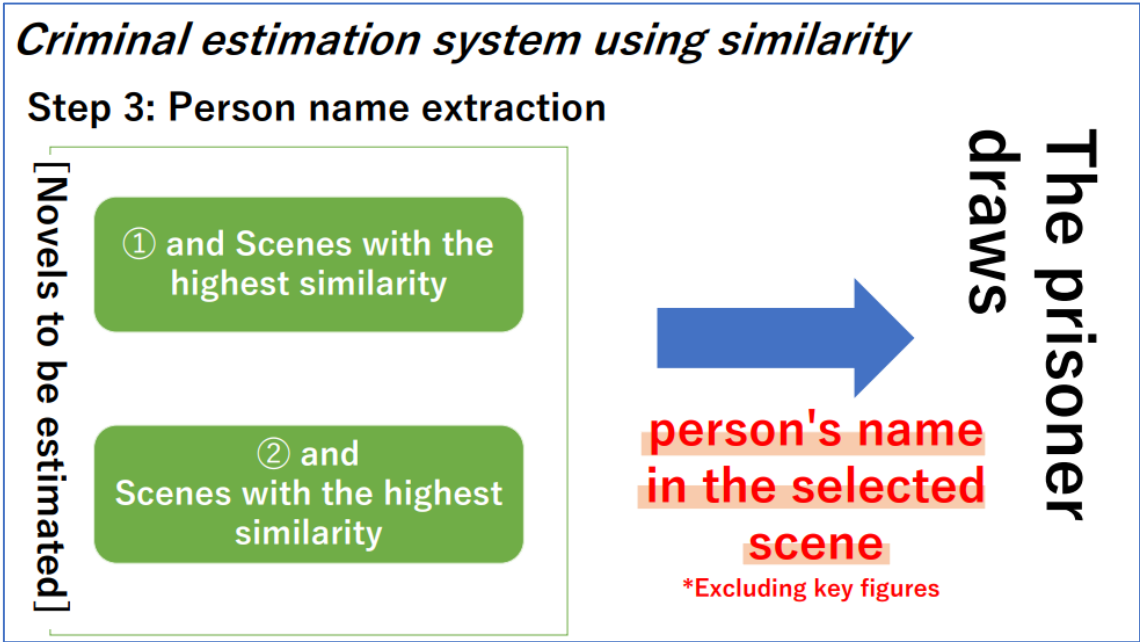
Scene Division

2.3 Similarity Calculation Illustration



Similarity Calculation

2.4 Extraction of culprits Illustration



Extraction of culprits

- Performance Information

(machine specs, run time, memory, etc.)

CPU: Intel(R) Core(TM) i3-8130U CPU @ 2.20 GHz 2.21 GHz

Memory : 4 GB

OS : windows 10

Execution time : about 10 minutes

3. Developed Application

- Source code is not available due to lack of preparation for submission

- Tools used

Google Colaboratory (free)

4. Sharing of materials

[Application Sheet]

- Public availability:

(☒) Allow to publish

(☐) Do not allow to publish

- Publication format

(☒) Published on the IKGRC website

(☐) Publish on your own site and wish to be linked to our site.

-> URL*:

[Submitted application, source code, data, etc.]

- Public availability:

(☒) Allow to publish

(☐) Do not allow to publish

- Publication format

(☒) Published on the IKGRC website

(☐) Publish on your own site and wish to be linked to our site.

-> URL*:

* The authors may contact us with the URL of the publication website after the website has been opened.