A Method to Extract **Knowledge Explanation** Sentences from Conversations in Comics with Combination of **Contents and Expressions**

Yoko Nishihara, Kohei Matsuoka (Ritsumeikan Univ.) and Ryosuke Yamanishi (Kansai Univ.)

Contact: Dr. Yoko Nishihara (nisihara at fc.ritsumei.ac.jp) Conversations in comics as resources for obtaining new knowledge

OComics represent the pop-culture of Japan.

• The characteristic of those is that conversations of characters drive stories.



OPeople can obtain new knowledge from conversations in real lives.

OThe conversations in comics could be useful to acquire new knowledge.

Definition: Knowledge explanation sentence (KES) in a conversation

 The conversations in comics contain various utterances.
 O Some of them explain knowledge.

• We define the utterance sentence as a knowledge explanation sentence (KES).

O In the example, the sentences in the 3rd balloon are KESs. He/she will know the word ``paper'' also means an academic article.



Example of knowledge explanation sentences (KESs) in a conversation

If the KESs are extracted automatically, what can be realized?

OSupporting learning with comics

- People can obtain new knowledge by reading comics. It is very fun.
- OThey can choose genres that they want to learn from topics of comics.
 - Oif they want medical knowledge, they can choose medical topic comics.
 - OA list of comics for learning has been published. The list was made manually. (https://gakushumanga.jp/).

ホーム プロジェクトについて 活動アーカイブ ニュース お問い合わせ

| | | | | | | | | C | すべてのジ | ャンルを表示 |
|------|------|------|------|------|------|------|------|------|-------|--------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 文 | 生 | 芸 | 社 | 職 | 歴 | 戦 | 生 | 科 | ス | 多 |
| 学 | 命 | 術 | 슾 | 業 | 史 | 争 | 活 | 学 | ポ | 様 |
| | ٢ | | | | | | | | I. | 性 |
| | 世 | | | | | | | 学 | ッ | |
| | 界 | | | | | | | 習 | | |
| (13) | (20) | (18) | (24) | (26) | (21) | (13) | (17) | (16) | (14) | (18) |

11月10日 【連載更新】とっぴぃ豊島の選択「これも学習マンガだ!大人も学べる学習入門」



Purpose and assumption of this research

• We propose an extraction method of the KESs from characters' conversations in comics.

• We assume that the KESs include content words and expressions that are special for knowledge explanation.

• We verify the assumption by evaluation experiments.

Outline of the proposed method



Text of utterance sentences by characters

• The utterance sentences are extracted from speech balloons.

- A label is assigned to each utterance sentence to indicate whether or not it is a KES.
- The labelling is conducted manually.

Averaged word vector of content words in an utterance sentence



• The content words are nouns, verbs and adjectives.

- O fastText is used for embedding.
- The dimension is convoluted from 300 to 59 in order to make the size of the two vectors same.

Expression vector of an utterance sentence

• We assume that when knowledge is explained, people must use characteristic function words.

O Conjunctions, auxiliaries, and particles in Japanese

- The meanings of the function words should be related to knowledge explanation.
 - O``However" (conjunction) has a meaning of paradox.
 - O``May" (auxiliary) has a meaning of guess.
- We use 12 meanings of conjunctions, 16 meanings of auxiliaries, and 31 meanings of particles.

O In total, the 59 dimensions vector is obtained for the expression vector.

Training the combination vectors by SVM

• The averaged vector of **content words and the expression vector** are connected with being a **combination vector**.

OThe combination vectors are trained by SVM.

Combination vector



Evaluation experiments

• We evaluated the accuracies, precisions, and recalls in extracting KESs from conversation texts in comics.

O Used method

- 1. Proposed method (averaged word vector of content words + expression vector trained by SVM)
- 2. Averaged word vector of content words trained by SVM
- 3. Expression vector trained by SVM

O Used comic data

- O 5 comics
- O 4,611 sentences
- O 665 KESs

| Title | Genre | Sentences | Knowledge | Rate (%) |
|------------------------|-------------|-----------|-----------|----------|
| Dr. STONE 1 | Chemistry | 851 | 44 | 5.2 |
| Cells at Work! 1 | Science | 976 | 136 | 13.9 |
| Kounodori: Dr. Stork 4 | Medical | 1,392 | 185 | 13.3 |
| Math Girls 1 | Mathematics | 671 | 166 | 24.7 |
| Gakken Manga | Education | 721 | 134 | 18.6 |
| Total | - | 4,611 | 665 | 14.4 |
| Average | - | 922 | 133 | 15.1 |

Experimental results

| | Combination vector | | | Averaged word vector | | | Expression vector | | |
|------------------|--------------------|-----------|--------|----------------------|-----------|--------|-------------------|------------|--------|
| Comic | Accuracy | Precision | Recall | Accuracy | Precision | Recall | Accuracy | Precisioin | Recall |
| Dr. STONE 1 | 0.79 | 0.75 | 0.88 | 0.75 | 0.74 | 0.82 | 0.58 | 0.68 | 0.51 |
| Cells at Work! 1 | 0.83 | 0.80 | 0.89 | 0.82 | 0.80 | 0.85 | 0.64 | 0.74 | 0.43 |
| Kounodori 4 | 0.81 | 0.77 | 0.88 | 0.81 | 0.77 | 0.88 | 0.68 | 0.72 | 0.60 |
| Sugaku Girl 1 | 0.81 | 0.79 | 0.86 | 0.79 | 0.77 | 0.84 | 0.69 | 0.78 | 0.54 |
| Gakken Manga | 0.80 | 0.75 | 0.89 | 0.76 | 0.71 | 0.88 | 0.70 | 0.75 | 0.60 |
| Average | 0.80 | 0.77 | 0.88 | 0.78 | 0.75 | 0.85 | 0.65 | 0.73 | 0.51 |
| | | | | | | | | | |

OThe proposed method marked 0.80 accuracy, 0.77 precision, and 0.88 recall, respectively.

Discussion

- The KESs often contain technical terms (delivery, congenital rubella syndrome, vaccination, mother and child shelter, rubella, and womb).
- The averaged vector of content words could model the characteristic of technical terms, so obtained higher indices.
- The proposed method marked the highest indices. It means the combination is necessary to extract the KESs.

Examples of extracted KESs.

分娩誘発なんかしたらさ患者死んじゃうよ
(If you try to induce a delivery, the patient will die.)
先天性風疹症候群って診断された赤ちゃんが 20 人超えたって…
(More than 20 babies diagnosed with congenital rubella syndrome...)
副作用などでワクチン接種が敬遠されて接種率がとても低い
(Vaccination rates are very low due to side effects and other factors that discourage vaccination.)
母子シェルターはご主人に居場所がわからないように他県になると思うの
(I think the mother and child shelter will be in another province so that your husband won't know where you are.)
私はね…お母さんのお腹の中で風疹にかかっちゃって

(I've... caught the rubella in my mom's womb.)

| | Combination vector | | | Averaged word vector | | | Expression vector | | |
|------------------|--------------------|-----------|--------|----------------------|-----------|--------|-------------------|------------|--------|
| Comic | Accuracy | Precision | Recall | Accuracy | Precision | Recall | Accuracy | Precisioin | Recall |
| Dr. STONE 1 | 0.79 | 0.75 | 0.88 | 0.75 | 0.74 | 0.82 | 0.58 | 0.68 | 0.51 |
| Cells at Work! 1 | 0.83 | 0.80 | 0.89 | 0.82 | 0.80 | 0.85 | 0.64 | 0.74 | 0.43 |
| Kounodori 4 | 0.81 | 0.77 | 0.88 | 0.81 | 0.77 | 0.88 | 0.68 | 0.72 | 0.60 |
| Sugaku Girl 1 | 0.81 | 0.79 | 0.86 | 0.79 | 0.77 | 0.84 | 0.69 | 0.78 | 0.54 |
| Gakken Manga | 0.80 | 0.75 | 0.89 | 0.76 | 0.71 | 0.00 | 0.70 | 0.75 | 0.60 |
| Average | 0.80 | 0.77 | 0.88 | 0.78 | 0.75 | 0.85 | 0.65 | 0.73 | 0.51 |
| | | | | | | | | | |

Conclusions

• We propose an extraction method of knowledge explanation sentences from conversations in comics.

• The method represents each utterance as a combination vector of content words and expressions.

• Experimental results showed that the proposed method obtained 0.80 accuracy, 0.77 precision, and 0.88 recall, respectively.

• As the future work, we try to use the extracted KESs for recommending comics and supporting learning with comics.





Procedures of the proposed method

- 1. Given a text of utterance sentence by characters, the method divides each sentence into words.
- 2. Content words are extracted and represented as embedding vectors. Then, all embedding vectors are averaged to obtain an **averaged word vector** of an utterance sentence.
- 3. Function words are extracted to make an **expression vector** of each utterance sentence.
- 4. The **two vectors** are combined with being a **combination vector**.
- 5. The **combination vectors** are trained by SVM to extract the KESs.

Making datasets for learning

• The number of KESs were 665 while that of non-KESs were 4,611.

O In order to prevent from overfitting, we made datasets that have almost the same numbers of KESs and non-KESs.

