Method for Automatically Generating Networks of Personal Relationships from Story Summaries

Jun Goto[1,2] **Masahiro Shibata**[1] **Nobuyuki Yagi**[1] **Akiko Aizawa**[2] **Satoshi Sekine**[3]

[1]NHK Science and Technical Research Laboratories
1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan

[2]The Graduate University for advanced studies / National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

[3] New York University
715 Broadway, 7th floor, New York, NY 10003, USA

ABSTRACT
We propose a method of automatically generating networks of personal relationships from program summaries, to enable an understanding of relations between story characters or persons in the programs. The method uses natural language processing to extract representations of the characters or persons from the target text. Then it creates relationships between these story characters or persons from correspondences found through syntax analysis. It determines whether anaphora relations exist between the representations, and it integrates the anaphora relations that refer to characters or persons to generate a network of personal relationships. The results of an evaluation revealed that the networks had F-measures of 0.907 for representation, 0.708 for relationship, and 0.800 for anaphora resolution. This shows that it is possible to extract relationships that cannot be extracted by using simply the co-occurrence frequency of personal names and that networks of personal relationships can be generated from short texts such as program summaries.

Author Keywords
Named entity recognition relation extraction, anaphora resolution, Personal Relationship

ACM Classification Keywords
I.2.7 [Natural Language Processing]: Text analysis

1. INTRODUCTION
Digital broadcasting provides metadata such as title, genre, and program summary along with each program. The summaries of movies and serial dramas include text metadata describing the background of the story, an outline of the events in the story, and information about the characters. Such information is helpful when viewers are selecting a program to watch. However, it is troublesome to have to read one summary after another when searching for a program to view. For people to comprehend the content of a story, it is necessary to understand the characters and the relations between them. If were possible for a machine to understand the content and display it in an easy-to-understand form, it would reduce the hassle endured by the viewer.

There has been research on extracting human relationships from web pages and depicting them as networks [1][2]. The research uses human co-occurrence frequencies for determining whether or not there are relationships, and it relies on a large volume of Internet data in which personal names appear frequently. Thus, if the subject corpus is smaller than a certain size, the frequency of people who are co-occurrent becomes sparse and it becomes difficult to separate out significant relations. There have also been attempts to create methods of classifying by finding co-occurrences of specific words, but it is not possible to classify relations that have not been envisioned beforehand.

Our study uses natural language processing, e.g., named entity recognition, relation extraction and anaphora resolution, to identify relationships between characters accurately in short passages of text such as program summaries, without the need for large-scale corpora. We integrate these relationships to generate a network in which characters in the story and people are represented as nodes and their relationships as edges. The aim is to develop an interface that helps users comprehend the content of program by visualizing the network of human relationships.

2. PROPOSED METHOD
The sequence for generating a network is shown in Figure 1. (1) From the program summary, we extract parts that deal with people. (2) We identify human relationships from the results of step (1) by performing syntax analysis. (3) We determine whether the representations refer to the same people by performing anaphora resolution. (4) From the results of step (2) and (3), we generate a network by integrating the identified relationships.

2.1 Human representation extraction
It is first necessary to extract the human representations appearing in that text. In the previous research [1][2], the generated network is based on the condition that personal names have been provided. With movies and dramas, however, it is not feasible to prepare the names of the characters in advance, since the characters and persons in them may

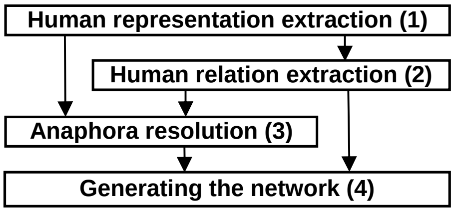

Figure 1. Flow of processing

Romantic comedy in which **B. Bardot** plays the role of a **young woman** called **Brigitte**, who is the **only daughter** of the **Prime Minister** of France. Madly in love with **Michel**, a **Cabinet Secretary**, **Brigitte** was initially snubbed but with a cunning ploy succeeded in a splendid marriage with **Michel**. However, disgusted by **his** brazen affairs, **she** announces she is having an affair with a **prince** who is **the nation's guest**. The role of the **prince** is played by the French **heartthrob Charles Boyer**, who was nominated for four Oscars as Best Actor in a Leading Role.

Figure 2. An example of TV program summary

change from episode to episode. Thus, our method automatically extracts representations from the program summaries. The representations are described below.

Named entities

Named entities are proper nouns such as personal names and numerical representations. Although personal names are most often used in text articles to represent persons, titles or occupations are also used instead of personal names in the summaries of programs. In particular, titles and occupations are often used as the representations of supporting characters. In this study, we chose personal names (Person) and titles/occupations (Title) as the possible human representations.

General human representations

In some cases, general nouns such as "woman" or "person" are used as representations, in which case no personal names appear. For that reason, we use these words as general human representations (G-Person). General human representations can include representations that indicate both a person and a relationship with another person, e.g., "only daughter." We extract these representations (R-Person) as well.

Pronouns

In written Japanese, personal pronouns such as "he" or "she" are rarely used; once a person's name has been introduced, that name is used continuously or is omitted. However, personal pronouns are used comparatively often in program summaries of movies and serial dramas. Our investigation of the summaries of 378 movie programs showed that personal pronouns were used in 158 (41.8%) of them. This result emphasizes that to grasp the stories, it is essential to extract the personal pronouns appearing in the program summaries.

Named entities are represented in a variety of ways or by integrating a number of morphemes. In addition, compound words are used in G- and R-Person representations. Machine learning (Conditional Random Fields) is therefore used to extract such representations. We used 57 features, including words, readings, parts of speech, character types, character N-gram (where N = 1, 2, 3) before and after each

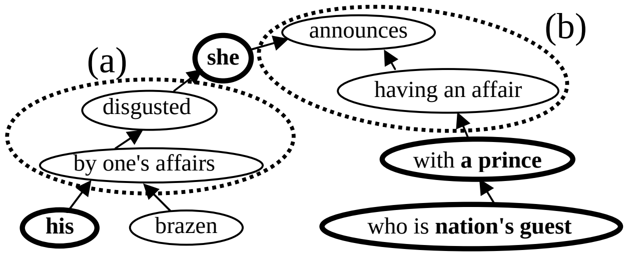

Figure 3. Relation extraction based on syntax analysis

word, inflections, and the broader concepts of Japanese concept dictionary EDR. Since there is a limited number of pronouns, we used parts of speech for extraction.

2.2 Human relation extraction

Relation extraction based on syntax analysis

To extract human relationships from small volumes of text, it is necessary to ascertain the structure of each sentence making up the text. A program summary has to introduce the characters and the relationships between them simply and within a limited space. For that reason, human representations often appear within one sentence. We use this characteristic to identify human relationships from single sentences.

Since Japanese has a different word order from English, it often happens that no relation appears between the positions of two human representations. We therefore used syntax analysis[1] to acquire the result of syntactic dependencies (a syntax tree), and used the syntax tree and the results of the representation extraction step to extract the relationships.

The syntax tree of the underlined text of Figure 2 is shown by way of example in Figure 3. We first take the correspondence between each human representation and the nodes of the syntax tree. We then acquire a representation between two nodes that include human representations on the syntax tree as a relation. As a result, representation (a), which is between the nodes "his" and "she," and representation (b), which is between "she" and "a prince," are extracted as relationship representations

Preprocessing of syntax analysis

The accuracy of the syntax analysis should be improved because the results of the relation extraction depend on it. Long sentences such as complex sentences tend to cause mistakes in the results. We tried to improve the accuracy of syntax analysis by dividing complex sentences into simple sentences. We used a pattern-based method [3] to divide them.

In Japanese, the subject or center of a sentence is often omitted from subsequent sentences. With a complex sentence, the subject is often omitted from the sentence. This omission is called a "zero pronoun," and it can be supplemented from the subject or topic of the previous sentence. Since relations are acquired between two or more human representations within one sentence, an omission would

[1] http://chasen.org/~taku/software/cabocha/

make it impossible to extract any relation. For a sentence in which the subject or topic is omitted, it is therefore necessary to obtain a plausible noun from a previous sentence to supplement the zero pronoun. For subject supplementation, we used an ordering that utilizes the surface layer case based on centering theory [4]. The priority ordering of the surface representation is as follows:

topic (“wa” clause) > subject (“ga” clause) > indirect object (“ni” clause) > direct object (“wo” clause) > others

2.3 Anaphora resolution

To generate a network, it is necessary to link together the relationships that have been obtained. Since not just personal names but also general nouns and pronouns are used to represent people in summaries, it is difficult to judge anaphoras from simply the similarity of character strings. This is because there are cases in which the same person is indicated even when there are different representations in the surface layer, and conversely, cases in which different people are indicated so no anaphoras are created, even though the representations are the same. For that reason, it is necessary to use a characteristic other than textual similarity to link together representations that indicate the same person.

If there is a contradiction between the results of a number of anaphors, it will be difficult with machine learning to represent a model that enables dynamic processing to resolve such contradictions. For that reason, our anaphora resolution is rules-based. It should be noted, however, that machine learning is used to extract the features that are used in the conditions within these rules. The types of anaphora resolution that we used are described below.

Pronominal anaphora

Anaphora resolution consists of processing that specifies an anaphor and antecedent. An anaphor is a pronoun (he, she, etc.) and a human representation continuing on from an adnominal (that man, that person, etc.). The determination of the antecedent starts from the sentence immediately before the anaphor, and the anaphoric sequencing is determined on the basis of an ordering that employs surface cases based on centering theory [4]. When there is no human representation in a candidate sentence, an antecedent from a previous sentence is nominated. Constraints are set on the selection of the antecedent by gender and singular/plural features. The determination of gender (male, female, or undefined) is done by multi-class SVMs, using the surface layer of the human representation and the reading as features. Singular/plural features are determined by using a dictionary.

Anaphora based on syntactic dependency

When a number of human representations are contained in the same sentence fragment, they are taken as referring to the same person or characters. Taking “only daughter Brigitte” as an example, “only daughter” and “Brigitte” refer to the same person. In addition, when representation A is connected to another representation B by a “no” (a possessive

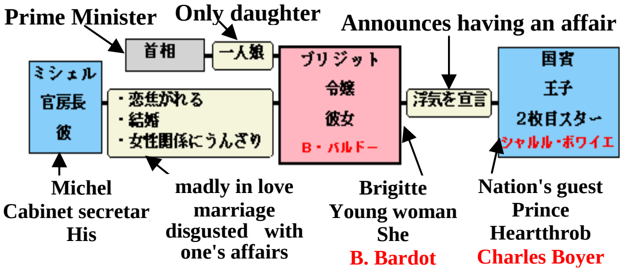

Figure 4. Generated network

in Japanese) clause, they are determined to be the same person or character.

Anaphora based on specific relations

Of the pairs of representations obtained by the relation extraction step that have some sort of relation, those that conform to previously defined relations are taken to be anaphoric relations. Twenty-eight types of relationship representation that were manually collected were handled as anaphoras. For example, there are phrases such as: “A is called B.” Words such as “perform” are acquired as anaphoras with a wide range of meanings concerning relationships between actors and character names.

Anaphora based on text similarity

In a similar manner to previous research, anaphors are determined by the similarity of strings. It should be noted, however, that if there is no consistency with other conditions, no anaphor is created even though the surfaces match. For example, when “detective A” and “detective B” appear, an anaphor result is created on the basis of syntactic dependency, but the personal names A and B create a contradiction so they do not make an anaphora even though the two instances of “detective” match.

2.4 Generating the network

We use the results of the human relation extraction and anaphora resolution to generate a network. This is shown as a graph with labels, where human representations in the anaphora relations are integrated as nodes and the relationships are shown as edges. The network created from the text of Figure 2 is shown in Figure 4. These results were used to define the network by using format dot in Graphviz[2], and they were drawn by using dedicated software. The colors of the nodes differentiate between the genders of people. In addition, the names of actors such as B. Bardot and Charles Boyer are displayed in red.

3. EVALUATION

To assess the capabilities of the system, we generated and evaluated networks. The development data included summaries of 207 movie programs and eight dramatic serials. For the evaluation data, we prepared 90 summaries that were not included in the development data. To obtain answers for the evaluation, we tagged the representations manually. We also tagged the relationships and the anaphora relations.

---

[2] http://www.graphviz.org/

Table 1 Results of human representation extraction

	Precision	Recall	F-measure
Person	0.940(202/215)	0.918(202/220)	0.929
Title	0.855(141/165)	0.834(141/169)	0.844
G-Person	0.842(117/139)	0.848(117/138)	0.845
R-Person	0.975(119/122)	0.992(119/120)	0.983
Pronoun	1.000(62/62)	0.984(62/63)	0.992
Total	0.912(641/703)	0.903(641/710)	0.907

Table 2 Results of relation extraction

Precision	Recall	F-measure
0.722 (179/248)	0.694 (179/258)	0.708

Table 3 Results of anaphora resolution

Type	Precision	Recall	F-measure
Syntactic-dependency	0.857 (60/70)	0.732 (60/82)	0.789
Text-similarity	0.955 (64/67)	0.821 (64/78)	0.883
Specific-relation	0.917 (33/36)	0.600 (33/55)	0.725
Pronominal	0.845 (49/58)	0.710 (49/69)	0.772
Total	0.892 (206/231)	0.725 (206/284)	0.800

3.1 Evaluation of human representation extraction

To verify whether the people in the generated networks were correct, we evaluated the representation step of our method. The results are shown in Table 1. The precision was 0.912, recall was 0.903, and F-measure was 0.907, even though the learning data was insufficient. The results of individual verification showed that the F-measures of G-Person and Title were 0.845 and 0.844, respectively. These results were low in comparison with the other representations. This is thought to be due to similarities between words that appear before and after target words, which led to many misclassifications between them, and to mismatches between classifications in EDR dictionary and the vocabulary appearing in the test data, which meant that the effects of semantic features were not sufficient. Regarding Title, the vocabulary spans a wide range of branches, so we believe that improvements could be achieved by increasing the amount of learning data.

3.2 Evaluation of relation extraction

Next, we verified the correctness of the generated relationships. The relationships were taken to be correct if the verbs and nouns that formed the major part of the text created by the system matched those of manual extraction. In this case, precision was 0.722, recall was 0.694, and F-measure was 0.708 (see Table 2). One reason for the drop in recall was that 9.7% of the representations could not be extracted in the representation step. Another reason was the lack of patterns for dividing up complex sentences; the syntax analysis failed in some cases when complex sentences could not be divided.

There were some examples in which the zero pronoun supplementation failed so that erroneous relations were extracted. This was because only case information based on centering theory was used in the supplementation. To resolve this problem, it will be necessary to introduce predicates that incorporate human representations and the associated case into the features. Other than the problem of the zero subject, there were cases in which the possessive case or object case was omitted, so it will be necessary to supplement zero pronouns overall to improve the accuracy of the relationship extraction.

3.3 Evaluation of anaphora resolution

Representations that form an anaphor are linked to one node in the network. To verify the accuracy of such links, we evaluated the anaphora resolution step. The results are shown in Table 3. Precision was 0.892, recall was 0.725, and F-measure was 0.800. The recall for Specific-relation was low in comparison with the other types of anaphora. The likely reason is that the scale of the development data was not sufficient so it was not possible to have exhaustive knowledge of all the Specific-relations. It is necessary to acquire more examples of Specific-relations from large-scale corpora such as newspapers and web pages, and use them as prior knowledge. Regarding Pronominal, there were many examples in which an antecedent word that was close in meaning ought to be taken, even with case ordering based on centering theory. To resolve problems of this type, we must set references for determining semantic information such as synonyms and superior/inferior concepts.

4. CONCLUSIONS

This paper described a method of extracting a network of human relationships from a program summary by using human representation extraction, human relation extraction and anaphora resolution. We evaluated this method, and the resulting F-measures were 0.907 for extracted representations, 0.708 for the extracted relationships, and 0.800 for anaphora resolution. By extracting pronouns and general nouns other than human names as human representations and by limiting anaphora resolution to human representations, we were able to generate networks of personal relationships that could not be created by conventional methods that simply use the co-occurrence frequencies of personal names.

REFERENCES

1. P. Mika: Flink: Semantic web technology for the extraction and analysis of social networks, Journal of Web Semantics, vol. 3, no. 2, pp. 211-223 (2005).

2. Y. Matsuo, J. Mori, M. Hamasaki, K. Ishida, T. Nishimura, H. Takeda, K. Hasida, and M. Ishizuka: POLYPHONET: An Advanced Social Network Extraction System from the Web, in Proc. of WWW-06, pp. 397–406 (2006).

3. Y. Kim and T. Ehara: A Method of Partitioning of Long Japanese Sentences with Subject Resolution in J/E Machine Translation, Proc. of International Conference on Computer Processing of Oriental Languages, pp. 467-473 (1994).

4. M. Walker, M. Iida, and S. Cote: Japanese discourse and the process of centering, Computational Linguistics, Vol. 20, No. 2, pp. 193-233 (1994).