知能化工学大部門　知覚情報処理研究分野
（高村大也研究室）

Email: takamura@pi.titech.ac.jp　　http://www.lr.pi.titech.ac.jp/~takamura/

（研究分野）

高村研究室では，計算言語学と自然言語処理の研究をしています。

（研究テーマ）

１）文書要約のための数理モデルの開発（高村大也）

文書が大量にあり，それらをすべて読むのが困難であるという状況が少なからず起こる。機械処理によりその内容を自動的に要約する技術が求められている。そこで，文書から文を選択して要約を作成する手法が盛んに研究されている（図１）。本研究室では，文の選択方法についての研究を行っている。我々は，施設配置問題と呼ばれる最適化問題によって文書要約をモデル化した（図４）。

このモデルは，推論関係や例示関係などの文間の関係を，自然な形で要約に活用できるという特長を持つ。また，これ以外にも最大被覆問題での文書要約のモデル化にも成功している。

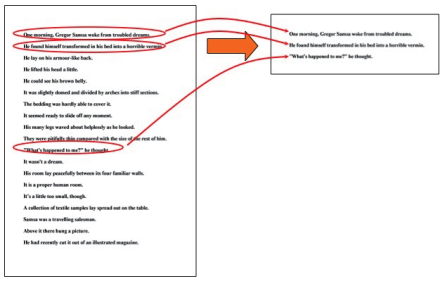
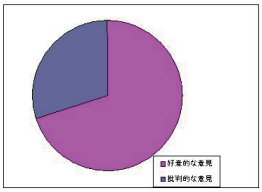


図１　文選択による要約生成

Fig.1　Summarization by sentence extraction

２）文書における感情・意見の解析技術（高村大也）

blogや電子掲示板といった新たな電子メディアの登場により，個人が手軽に自分の意見を発信できるようになった。この個人の意見は，さまざまな状況で有益である。例えば，携帯電話の購入を検討している時に，「Ｘ社の携帯は壊れにくい」，「Ｙ社の画面は文字が見えづらい」というような意見を横並びで閲覧できれば，より迅速な意思決定ができるだろう．このような背景のもと，本研究室では，電子文書から人々の意見や感情を自動的に抽出，整理する研究を進めている（図２）。また，「日本代表は４位に終わった」を「日本代表は４位に入賞した」と変換するなど，意味的には同内容だが良い印象を与える表現への言い換え技術についても研究を進めている（図５）。



好意的意見の例：
「電池が長持ち」、「軽くて持ち運びやすい」、「安い」

批判的意見の例：
「すぐに壊れる」、「使い方が難しい」、「高い」

図２　携帯音楽プレーヤに関する評判の自動解析例

Fig.2　Summary of reviews on handy music players

３）質問応答サイトにおける発言間の関係同定とそれを用いた要約（神保一樹，高村大也）

インターネット上の質問応答サイトには，膨大な知識が眠っている。この知識を有効利用したいという要望がある。しかし，図３に示すように，一つの質問に対して多くの類似した回答がポストされるなどしており，それらをすべて読むことは効率的でない。そこで，質問応答を要約する技術を開発した。開発手法においては，まず質問応答サイトの発言（質問及び回答）間の関係（例えば類似関係や包含関係など）を特定する（図６）。これらの特定した関係を用いて，なるべく情報が冗長にならないように発言を選択することで，質問応答の要約を実現している。

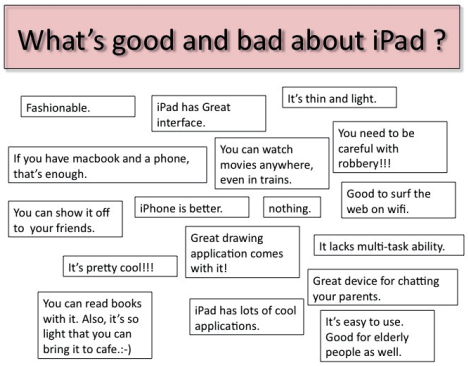


図３　WEB上の質問応答サイトにおける質問と回答の例

Fig.3　An example of a question and answers on a QA site on the web

Advanced Information Processing Division
Intelligent Information Processing Section
(Hiroya Takamura Group)

(Research Field)

Computational linguistics and natural language processing

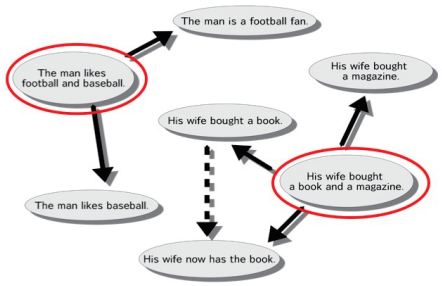


図４　施設配置問題による文書要約モデル

Fig.4　Text summarization model based on facility location problem

「日本は4位に終わった」→「日本は4位入賞を果たした」
「彼はでしゃばりだ」→「彼にはリーダーシップがある」
「しゃべりすぎでうるさい」→「話題が豊富で周りを飽きさせない」
「自分勝手だ」→「他人に流されぬ強い意志を持つ」
「古臭い建物だ」→「歴史を感じさせる建造物である」
「It's sort of arty.」→「It's quite artistic.」

図５　ポジティブ表現への言い換え例

Fig.5　Paraphrases to positive expressions

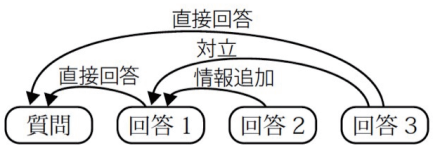


図６　発言間の関係例

Fig.6　An example of relations between utterances

(Current Topics)

1) Mathematical models for text summarization

(Hiroya Takamura)

People often come across the situation where there is too much text to read. In such a situation, a technique that automatically summarizes the entire text will be very useful. Many researchers are working on the text summarization methods, many of which are based on sentence extraction (Fig.1). We developed a novel text summarization method. Our summarization model is based on the facility location problem, which is a discrete optimization problem. This model can make a good use of relations between sentences such as entailment relation and exemplification relation.

2) Analysis of Emotions and Opinions in Text Data

(Hiroya Takamura)

With the advent of blogs and web BBSs, people can easily express our own opinions today. Such numerous opinions on the web are often useful. For example, when we are going to buy a new handy phone, we can get many anonymous advices on the web, such as 'Mobile phones made by X company get easily broken.' Thus we are working on automatic extraction or analysis methods for opinions and emotions in text (Fig. 2). We are also working on paraphrasing techniques from a negative expression to its positive counterpart, such as "This chair is a little arty." to "This chair is artistic" (Fig.5).

3) Relation identification of utterances on QA sites and the applications to summarization

(Kazuki Jinbo, Hiroya Takamura)

There is a huge amount of knowledge on QA sites on the internet, in which users ask questions and other users answer those questions. However, one question can have many redundant answers as in Fig. 3, and it is sometimes very inefficient to read them all. We hence developed a technique for summarizing question and answers. In our technique, we first identify the relations between utterances (i.e., question or answers), such as equivalence as in Fig. 6. We then use the identified relations to generate a summary that has as few redundant utterances as possible.