

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-318509  
(P2006-318509A)

(43) 公開日 平成18年11月24日(2006.11.24)

(51) Int. Cl.	F I	テーマコード (参考)
<b>G06F 17/30 (2006.01)</b>	G06F 17/30 220A	5B009
<b>G06F 17/21 (2006.01)</b>	G06F 17/30 170A	5B075
	G06F 17/21 550A	

審査請求 有 請求項の数 6 O L (全 26 頁)

(21) 出願番号	特願2006-222723 (P2006-222723)	(71) 出願人	301022471
(22) 出願日	平成18年8月17日 (2006.8.17)		独立行政法人情報通信研究機構
(62) 分割の表示	特願2004-168944 (P2004-168944) の分割		東京都小金井市貫井北町4-2-1
原出願日	平成14年2月22日 (2002.2.22)	(74) 代理人	100119161
			弁理士 重久 啓子
		(74) 代理人	100111822
			弁理士 渡部 章彦
		(72) 発明者	村田 真樹
			東京都小金井市貫井北町4-2-1 独立行政法人情報通信研究機構内
		Fターム(参考)	5B009 QA05 5B075 ND03 NS01

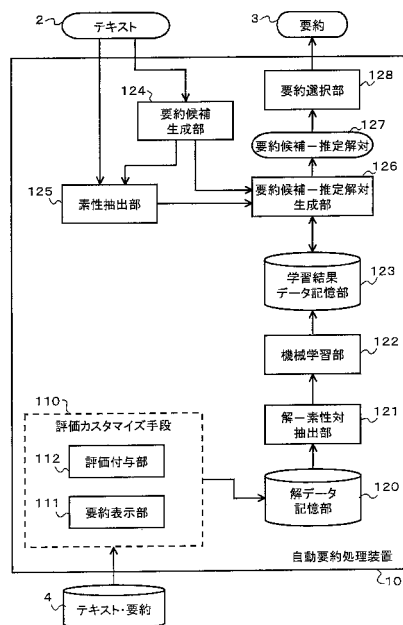
(54) 【発明の名称】 解データ編集処理装置および処理方法

(57) 【要約】

【課題】 機械学習法による要約処理で用いる解データをユーザが自由に編集できる編集処理と前記解データを用いてユーザに特化した要約処理とを実現する。

【解決手段】 評価カスタマイズ手段110は、テキスト・要約4からユーザが指定した範囲の文データを抽出してユーザ指定要約とし、複数の要約性質ごとの評価の入力項目を表示し、ユーザ指定要約についてユーザの評価入力を受け付け、所定規則または機械学習法による自動要約生成処理によりテキストの要約を生成する。そして、テキストとユーザ指定要約で構成される問題に対してユーザ入力評価である解を付与したデータと、テキストとユーザ指定要約以外の部分で構成される問題に対してユーザ指定要約ではないことを示す悪い評価である解を付与したデータとを含む解データを生成して解データ記憶部120に出力する。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

文書データであるテキストを機械学習法により自動要約する処理で用いる解データを編集する解データ編集処理装置であって、

文書データであるテキストを記憶するテキスト記憶手段と、

前記テキスト記憶手段から取得したテキストを表示装置に表示する処理手段と、

前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とし、複数の要約の性質ごとの評価を入力する項目を表示して、前記ユーザ指定要約について前記性質ごとのユーザの評価の入力を受け付ける処理手段と、

所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成する処理手段と、

前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記性質ごとに前記ユーザが入力した評価を解として付与したデータと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記性質ごとにユーザ指定要約ではないことを示す所定の悪い評価を解として付与したデータとを含む解データを生成して解データ記憶手段に出力する処理手段とを備える

ことを特徴とする解データ編集処理装置。

## 【請求項 2】

請求項 1 に記載の解データ編集処理装置において、

前記要約を生成する処理手段は、さらに、ユーザによって指定された部分の語句の変更の入力を受け付け、変更された前記部分を前記ユーザ指定要約とする

ことを特徴とする解データ編集処理装置。

## 【請求項 3】

請求項 1 に記載の解データ編集処理装置において、

前記要約を生成する処理手段は、前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるものに対して前記性質ごとの評価を入力する項目を表示し、前記要約について前記性質ごとのユーザの評価の入力を受け付け、

前記解データを出力する処理手段は、前記自動要約生成処理による要約の評価としてユーザの評価の入力を受け付けた場合に、前記テキストおよび前記要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記要約に対してユーザが入力した評価を解として付与したデータを解データとして生成する

ことを特徴とする解データ編集処理装置。

## 【請求項 4】

コンピュータが、文書データであるテキストを機械学習法により自動要約する処理で用いる解データを編集する解データ編集処理方法であって、

文書データであるテキストを記憶するテキスト記憶手段にアクセスしてテキストを取得する処理過程と、

前記テキストを表示装置に表示する処理過程と、

前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とし、複数の要約の性質ごとの評価を入力する項目を表示して、前記ユーザ指定要約について前記性質ごとのユーザの評価の入力を受け付ける処理過程と、

所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成する処理過程と、

前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記性質ごとに前記ユーザが入力した評価を解として付与したデータと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記性質ごとにユーザ指定要約ではないことを示す所定の悪い評価を解として付与したデータとを含む解データを生成して解データ記憶手段に出力する処理過程とを備える

ことを特徴とする解データ編集処理方法。

## 【請求項 5】

請求項 4 に記載の解データ編集処理方法において、

前記要約を生成する処理過程では、さらに、ユーザによって指定された部分の語句の変更の入力を受け付け、変更された前記部分を前記ユーザ指定要約とすることを特徴とする解データ編集処理方法。

【請求項 6】

請求項 4 に記載の解データ編集処理方法において、

前記要約を生成する処理過程では、前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるものに対して前記性質ごとの評価を入力する項目を表示し、前記要約について前記性質ごとのユーザの評価の入力を受け付ける処理を行い、

前記解データを出力する処理過程では、前記自動要約生成処理による要約の評価としてユーザの評価の入力を受け付けた場合に、前記テキストおよび前記要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記要約に対してユーザが入力した評価を解として付与したデータを解データとして生成する処理を行う

10

ことを特徴とする解データ編集処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、機械学習法により文章を自動要約する処理において、編集可能な解データの編集処理および解データを用いる機械学習法を用いた自動要約処理に関する。

【背景技術】

20

【0002】

近年、情報技術の発展に伴ってコンピュータを用いた文章の自動要約処理が盛んになってきている。しかし、個人的な嗜好や要約結果の用途などにより、所望する要約結果の傾向に相違があると考えられる。

【0003】

例えば、以下の非特許文献 1 では、複数の者がそれぞれ重要文抽出による要約を行なった結果に対する相互評価の尺度として再現率と適合率とを求めて表 4 に示している。非特許文献 1 の表 4 から明らかなように、20 文を抽出する処理の場合に、人 - 人（評価者相互）の評価（再現率および適合率）は、各評価者 A, B, C の一致度は 50 ~ 70 % であってあまり高い値とはいえず、要約結果に対する評価に個人差が存在することが推定できる。

30

【0004】

また、以下の非特許文献 2 では、サポート・ベクトル・マシン（Support Vector Machine）による重要文抽出処理において、処理セット A, B, C について交差検定の精度が最もよいことを表 4 により示している。非特許文献 2 の表 4 に示された交差検定は、同一評価者による処理と同一視でき、セット A, B, C を作成した者が同一かどうかは不明であるが、少なくとも同一時期もしくは同一人物により学習データを作成したほうが精度が良いということがわかる。

【非特許文献 1】伊藤山彦他、「講演文を対象にした重要文抽出」、言語処理学会第 7 回年次大会発表論文集、言語処理学会、2001年、pp.305-308

40

【非特許文献 2】平尾勉他、「Support Vector Machineによる重要文抽出」、情報学会基礎論文 63-16、情報学会、2001年、pp.121-127

【発明の開示】

【発明が解決しようとする課題】

【0005】

このように、要約結果に対する評価に個人差や用途差が存在すると考えることができることから、機械学習法を用いた自動要約処理においても、同じ評価にもとづいた要約を行なうのではなく、ユーザに特化した要約ができる必要がある。そのために、教師となる解データをユーザが自由に編集できる必要がある。

【0006】

50

本発明の目的は、機械学習法で用いる解データとなる要約結果または要約結果に対する評価をユーザが任意に編集できる解データの編集処理を実現することである。および、この解データを用いた機械学習法を用いてユーザごとに特化した要約を行える自動要約処理技術を実現することである。

【課題を解決するための手段】

【0007】

上記の目的を達成するため、本発明は、予め備えた解データを用いた機械学習処理に対してユーザがどのような要約結果を高く評価したかの情報をフィードバックするために、ユーザが要約結果やその評価を編集できるようにする。また、ユーザが編集した解データのフィードバックにより、機械学習処理においてユーザごとの特性を学習し、ユーザに特化した要約を行なうことができるようにするものである。

10

【0008】

本発明は、文書データであるテキストを機械学習法により自動要約する処理で用いる解データを編集する解データ編集処理装置であって、1) 文書データであるテキストを記憶するテキスト記憶手段と、2) 前記テキスト記憶手段から取得したテキストを表示装置に表示する処理手段と、3) 前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とし、複数の要約の性質ごとの評価を入力する項目を表示して、前記ユーザ指定要約について前記性質ごとのユーザの評価の入力を受け付ける処理手段と、4) 所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成する処理手段と、5) 前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記性質ごとに前記ユーザが入力した評価を解として付与したデータと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記性質ごとにユーザ指定要約ではないことを示す所定の悪い評価を解として付与したデータとを含む解データを生成して解データ記憶手段に出力する処理手段とを備えることを特徴とする。

20

【0009】

また、本発明は、文書データであるテキストを機械学習法を用いて自動要約する自動要約処理装置であって、1) テキストおよび前記テキストの要約を記憶するテキスト記憶手段と、2) 前記要約を表示装置に表示する要約表示処理手段と、3) 前記要約に対するユーザの評価の入力を受け付けて前記要約の評価とする評価設定処理手段と、4) 前記テキストおよび前記要約で構成される問題に対し前記評価を解として付与して生成した解データを解データ記憶手段に記憶する解データ出力処理手段と、5) 前記解データから前記問題の素性の集合と前記解との組を抽出し、当該組から、どのような素性のときにどのような解となりやすいかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、6) 要約対象のテキストを入力し、前記入力テキストから要約候補を生成する要約候補生成処理手段と、7) 前記入力テキストおよび前記要約候補から素性の集合を抽出し、当該素性の集合からどのような解となりやすいかを前記学習結果データをもとに推定し、要約候補と推定解との対を生成する要約候補 - 推定解対生成処理手段と、8) 前記要約候補 - 推定解対から、推定解が所定の良い評価でかつ確信度が最高の対を選択し、当該対の要約候補を要約とする要約選択処理手段とを備える。

30

40

【0010】

本発明は、テキスト記憶手段に記憶されたテキストの要約を表示装置に表示し、前記要約に対するユーザの評価の入力を受け付けて前記要約の評価とする。そして、前記テキストおよび前記要約で構成される問題に対し前記評価を解として付与して生成した解データを解データ記憶手段に記憶し、前記解データから前記問題の素性の集合と前記解との組を抽出し、当該組から、どのような素性のときにどのような解となりやすいかを学習した学習結果データを学習結果データ記憶手段に記憶する。

【0011】

その後、要約対象のテキストを入力し、前記入力テキストから要約候補を生成し、前記

50

入力テキストおよび前記要約候補から素性の集合を抽出し、当該素性の集合からどのような解となりやすいかを前記学習結果データをもとに推定し、要約候補と推定解との対（要約候補 - 推定解対）を生成する。そして、前記要約候補 - 推定解対から、推定解が所定の良い評価でかつ確信度が最高の対を選択し、当該対の要約候補を要約とする。

【0012】

これにより、表示した要約に対するユーザの評価を用いてユーザが良いと考える要約を機械学習し、その後に入力したテキストについてユーザに特化した要約を行うことができる。

【0013】

または、本発明は、1) テキストを記憶するテキスト記憶手段と、2) 前記テキストを表示装置に表示するテキスト表示処理手段と、3) 前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とする要約編集処理手段と、4) 所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成し、前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記ユーザによって選ばれた良い要約であることを示す所定の良い評価を解として付与して生成した解データと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記ユーザ指定要約ではないことを示す所定の悪い評価を解として付与して生成した解データとを解データ記憶手段に記憶する解データ出力処理手段と、5) 前記解データから前記問題の素性の集合と前記解との組を抽出し、当該組から、どのような素性のときにどのような解となりやすいかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、6) 要約対象のテキストを入力し、前記入力テキストから要約候補を生成する要約候補生成処理手段と、7) 前記入力テキストおよび前記要約候補から素性の集合を抽出し、当該素性の集合からどのような解となりやすいかを前記学習結果データをもとに推定し、要約候補と推定解との対を生成する要約候補 - 推定解対生成処理手段と、8) 前記要約候補 - 推定解対から、推定解が所定の良い評価でかつ確信度が最高の対を選択し、当該対の要約候補を要約とする要約選択処理手段とを備える。

10

20

【0014】

本発明は、テキスト記憶手段に記憶されたテキストを表示装置に表示し、前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とする。そして、所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成し、前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記ユーザによって選ばれた良い要約であることを示す所定の良い評価を解として付与して生成した解データと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記ユーザ指定要約ではないことを示す所定の悪い評価を解として付与して生成した解データとを解データ記憶手段に記憶する。さらに、前記解データから前記問題の素性の集合と前記解との組を抽出し、当該組から、どのような素性のときにどのような解となりやすいかを学習した学習結果データを学習結果データ記憶手段に記憶する。その後、要約対象のテキストを入力し、前記入力テキストから要約候補を生成し、前記入力テキストおよび前記要約候補から素性の集合を抽出し、当該素性の集合からどのような解となりやすいかを前記学習結果データをもとに推定し、要約候補と推定解との対（要約候補 - 推定解対）を生成する。そして、前記要約候補 - 推定解対から、推定解が所定の良い評価でかつ確信度が最高の対を選択し、当該対の要約候補を要約とする。

30

40

【0015】

これにより、表示したテキストからユーザによって抽出された部分をユーザが良いと評価した要約として機械学習し、その後に入力したテキストについてユーザに特化した要約を行うことができる。

50

## 【0016】

または、本発明は、1) テキストおよび前記テキストの要約を記憶するテキスト記憶手段と、2) 前記要約を表示装置に表示する要約表示処理手段と、3) 前記要約に対するユーザの評価の入力を受け付けて前記要約の評価とする評価設定処理手段と、4) 前記テキストおよび前記要約で構成される問題に対し前記ユーザが設定した評価を解として付与した解データを生成し、解データ記憶手段に記憶する解データ出力処理手段と、5) 所定の評価のうち前記解となった評価以外の評価を解候補として、前記解データから前記問題の素性の集合と解もしくは解候補との組を抽出し、前記素性の集合と解との組を正例と前記素性の集合と解候補との組を負例とする素性・解対・素性・解候補対抽出処理手段と、6) 前記抽出した組を教師信号として、どのような解もしくは解候補と素性の集合のときに正例である確率または負例である確率となるかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、7) 要約対象のテキストを入力し、前記入力テキストから要約候補を生成する要約候補生成処理手段と、8) 前記所定の評価を解の候補として、前記入力テキストおよび前記要約候補から素性の集合と解の候補との組を生成し、当該素性の集合と解の候補の組の場合に正例もしくは負例である確率を前記学習結果データをもとに推定し、前記推定した結果を推定解として前記要約候補と解の候補の組と前記推定解との対を生成する要約候補・推定解対生成処理手段と、9) 前記要約候補と解の候補の組・推定解対から、解の候補が所定の良い評価でかつ推定解の正例の確率が最高の対を選択し、当該対の要約候補を要約とする要約選択処理手段とを備える。

10

## 【0017】

または、本発明は、1) テキストを記憶するテキスト記憶手段と、2) 前記テキストを表示装置に表示するテキスト表示処理手段と、3) 前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とする要約編集処理手段と、4) 所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成し、前記テキストおよび前記ユーザ指定要約で構成される問題に対し前記ユーザによって選ばれた良い要約であることを示す所定の良い評価を解として付与して生成した解データと、前記テキストおよび前記自動要約生成処理による要約であって前記ユーザ指定要約以外の部分からなるもので構成される問題に対し前記ユーザ指定要約ではないことを示す所定の悪い評価を解として付与して生成した解データとを解データ記憶手段に記憶する解データ出力処理手段と、5) 所定の評価のうち前記解となった評価以外の評価を解候補として、前記解データから前記問題の素性の集合と解もしくは解候補との組を抽出し、前記素性の集合と解との組を正例と前記素性の集合と解候補との組を負例とする素性・解対・素性・解候補対抽出処理手段と、6) 前記抽出した組を教師信号として、どのような解もしくは解候補と素性の集合のときに正例である確率または負例である確率となるかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、7) 要約対象のテキストを入力し、前記入力テキストから要約候補を生成する要約候補生成処理手段と、8) 前記所定の評価を解の候補として、前記入力テキストおよび前記要約候補から素性の集合と解の候補との組を生成し、当該素性の集合と解の候補の組の場合に正例もしくは負例である確率を前記学習結果データをもとに推定し、前記推定した結果を推定解として前記要約候補と解の候補との組と前記推定解との対を生成する要約候補・推定解対生成処理手段と、9) 前記要約候補と解の候補との組・推定解対から、解の候補が所定の良い評価でかつ推定解の正例の確率が最高の対を選択し、当該対の要約候補を要約とする要約選択処理手段とを備える。

20

30

40

## 【0018】

または、本発明は、1) テキストを記憶するテキスト記憶手段と、2) 前記テキストを表示装置に表示するテキスト表示処理手段と、3) 前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とする要約編集処理手段と、4) 前記テキストを問題とし前記問題に対する前記ユーザ指定要約を解とする解データを生成し解データ記憶手段に記憶する解データ出力処理手段と、5) 前記解データから

50

前記問題の素性の集合と前記解との組を抽出し、当該組から、どのような素性のときにどのような解となりやすいかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、6)要約対象のテキストを入力し、前記入力テキストから素性の集合を抽出し、当該素性の集合からどのような解となりやすいかを前記学習結果データをもとに推定する解推定処理手段と、7)前記解推定処理手段で推定された解を前記入力テキストの要約として出力する要約選択処理手段とを備える。

【0019】

これにより、表示した要約に対するユーザの評価を用いて要約処理を機械学習し、入力したテキストについてユーザに特化した要約を行うことができる。

【0020】

または、本発明は、1)テキストを記憶するテキスト記憶手段と、2)前記テキストを表示装置に表示するテキスト表示処理手段と、3)前記テキストからユーザによって指定された範囲の文データを抽出して前記テキストのユーザ指定要約とする要約編集処理手段と、4)前記テキストを問題とし前記問題に対する前記ユーザ指定要約を解とする解データを生成し解データ記憶手段に記憶する解データ出力処理手段と、5)所定の規則に基づいて要約を生成する処理もしくは機械学習法を用いて要約を生成する処理のいずれかの自動要約生成処理により前記テキストの要約を生成し、前記要約のうち前記ユーザ指定要約以外の部分からなるものを解候補とし、前記解データから解もしくは解候補と前記問題の素性の集合との組を抽出し、前記素性の集合と解との組を正例と前記素性の集合と解候補との組を負例とする素性-解対・素性-解候補対抽出処理手段と、6)前記抽出した組を教師信号として、どのような解もしくは解候補と素性の集合のときに正例である確率または負例である確率となるかを学習した学習結果データを学習結果データ記憶手段に記憶する機械学習処理手段と、7)要約対象のテキストを入力し、前記入力テキストから要約候補を生成する要約候補生成処理手段と、8)前記要約候補を解の候補として、前記入力テキストおよび前記要約候補から素性の集合と解の候補との組を生成し、当該素性の集合と解の候補との組の場合に正例もしくは負例である確率を前記学習結果データをもとに推定し、前記推定した結果を推定解として、前記要約候補と推定解との対を生成する要約候補-推定解対生成処理手段と、9)前記要約候補-推定解対から前記推定解の正例の確率が最高の対を選択し、当該対の要約候補を要約とする要約選択処理手段とを備える。

10

20

【0021】

これにより、表示したテキストから抽出された部分をユーザが良いと評価した要約として要約処理を機械学習し、入力したテキストについてユーザに特化した要約を行うことができる。

30

【0022】

本発明にかかる処理装置の各手段または機能または要素は、コンピュータが実行可能なプログラムによっても実現できる。このプログラムは、コンピュータが読み取り可能な、可搬媒体メモリ、半導体メモリ、ハードディスクなどの適当な記録媒体に格納することができ、これらの記録媒体に記録して提供され、または、通信インタフェースを介して種々の通信網を利用した送受信により提供されるものである。

【発明の効果】

40

【0023】

本発明によれば、ユーザは、機械学習の解データとされる要約結果に対する評価を任意に設定することができるため、コンピュータを用いた自動要約処理においても、一つの類型で要約するのではなく、ユーザに特化した要約を行なえることが可能となる。

【0024】

機械学習法を用いた自動要約処理においても、同じ評価にもとづいた要約を行なうのではなく、ユーザに特化した要約を可能にするために、教師となる解データをユーザが自由に編集できる。

【0025】

また、同一人物であっても要約の評価が変化することが考えられるが、本発明によれば

50

、同一人物であっても随時要約結果に対する評価を設定でき、新たな解データを用いて機械学習し直すことにより、新しい評価態度に合わせた要約を行なうことが可能となる。

【発明を実施するための最良の形態】

【0026】

〔第1の実施の形態〕

図1に、第1の実施の形態における本発明の処理装置の構成例を示す。

【0027】

自動要約処理装置10は、評価カスタマイズ手段110と、解データ記憶部120と、解-素性対抽出部121と、機械学習部122と、学習結果データ記憶部123と、要約候補生成部124と、素性抽出部125と、要約候補-推定解対生成部126と、要約選択部128とを備える。

10

【0028】

評価カスタマイズ手段110は、解データ編集処理を実現する処理手段である。また、解データ記憶部120と、解-素性対抽出部121と、機械学習部122と、学習結果データ記憶部123とは、特許請求の範囲に示す自動要約処理装置の機械学習処理手段を実現する処理手段である。

【0029】

評価カスタマイズ手段110は、要約結果やその評価をユーザごとにカスタマイズする手段であって、要約表示部111と、評価付与部112とを備える。

【0030】

要約表示部111は、予め用意されたテキスト・要約4の要約結果を表示装置(図1に図示しない)に表示する手段である。

20

【0031】

テキスト・要約4は、テキストとその要約結果からなる。テキストは、一または複数の記事などからなる文書データである。要約結果は、テキストを要約した文書データである。要約結果としては、人手で生成したもの、自動要約処理装置10が入力したテキスト2に対して出力した要約3もしくは要約候補生成部124が生成し解データ記憶部120に記憶した要約候補であってもよい。

【0032】

評価付与部112は、要約表示部111が表示した要約結果に対してユーザが入力した評価を付与し、または、要約結果に予め与えられている評価をユーザが入力した評価に変更する手段である。

30

【0033】

解データ記憶部120は、機械学習部122が機械学習法を実行する際に教師とする解データを記憶する手段である。解データ記憶部120には、解データとして、テキストおよびその要約結果とからなる問題と要約結果に対する評価である解との組である事例が記憶される。

【0034】

解-素性対抽出部121は、解データ記憶部120に記憶されている事例ごとに解と素性の集合との組を抽出する手段である。

40

【0035】

素性とは、解析に用いる情報の細かい1単位を意味し、ここでは、1)文のなめらかさを示す情報、2)内容をよく表しているかどうかを示す情報、および、3)自動要約処理で用いられる特徴的な情報などである。

【0036】

機械学習部122は、解-素性対抽出部121により抽出された解と素性の集合との組から、どのような素性の集合のときにどのような解になりやすいかを機械学習法により学習し、学習結果を学習結果データ記憶部123に保存する手段である。機械学習部122は、解データを用いた機械学習法であればどのような手法で処理を行ってもよい。手法としては、例えば、決定木法、サポートベクトル法、パラメータチューニング法、シンプル

50



ベイズ法，最大エントロピー法，決定リスト法などがある。

【0037】

学習結果データ記憶部123は，機械学習部122の学習結果データを記憶する手段である。

【0038】

要約候補生成部124は，入力されたテキスト2から，所定の方法にもとづいて要約候補を生成する手段である。要約候補生成部124は，重要文選択モデル，重要箇所選択モデル，変形規則を利用したモデル，ランダムジェネレーションを利用したモデルなどの種々のモデルを用いて要約候補を生成する。

【0039】

素性抽出部125は，テキスト2および要約候補生成部124で生成された要約候補について素性の集合を抽出して要約候補 - 推定解対生成部126へ渡す手段である。

【0040】

要約候補 - 推定解対生成部126は，学習結果データ記憶部123の学習結果データを参照して，素性抽出部125から渡された素性の集合の場合に，どのような解になりやすいかを推定して，要約候補と推定解との対（要約候補 - 推定解対）127を生成する手段である。要約候補 - 推定解対生成部126は，さらに，各要約候補 - 推定解対127に，その推定解である確信度（確率）を求めて付与しておく。

【0041】

要約選択部128は，要約候補 - 推定解対127を受け取り，確信度の値が最も高い要約候補 - 推定解対127を選択し，その要約候補を要約3とする手段である。

【0042】

第1の実施の形態における評価カスタマイズ処理を説明するため，3人のユーザA，B，Cが要約結果をカスタマイズする場合を考える。

【0043】

ユーザAは要約結果に精度に関する記載が含まれていることを重視して評価すると仮定する。ユーザBは要約結果に手法に関する記載が含まれていることを重視し，ユーザCは，要約結果に手法と精度の両方に関する記載が含まれていることを重視して評価すると仮定する。また，要約結果の評価を3段階に分けて，評価1 = よい，評価2 = どちらでもない，評価3 = 悪い，のいずれかの分類先（評価）を与えるとする。

【0044】

図2に，第1の実施の形態における評価カスタマイズ処理の流れを示す。

【0045】

まず，テキスト・要約4が用意されているとする。図3にテキスト・要約4のテキストの例を示し，図4に要約結果の例を示す。図4(A)～(C)のそれぞれに，3つの要約結果 $r_1$ ， $r_2$ ， $r_3$ を示す。

【0046】

要約表示部111は，テキスト・要約4から取り出した要約結果を表示画面に表示する（ステップS1）。そして，評価付与部112は，ユーザが入力した評価を受け付け，その入力された評価を表示された要約結果の解（評価）とする（ステップS2）。

【0047】

ここで，ユーザAが自動要約処理装置10を使用する場合を想定する。ユーザAは，図4(A)の要約結果 $r_1$ に対して，精度に関係することが要約結果として抽出されているため，評価1をつける。すると，評価付与部112は，ユーザの入力（評価1）を受け付けて，事例c1の解として評価1を設定する。

【0048】

次に，要約表示部111が図4(B)に示す事例c2の要約結果 $r_2$ を表示した場合には，要約結果 $r_2$ は精度に関係することが抽出されていないため，ユーザAは，要約結果 $r_2$ に対して評価3をつけ，評価付与部112は，事例c2の解として評価3を設定する。

。

10

20

30

40

50

## 【0049】

さらに、要約表示部 111 が図 4 (C) に示す事例 c3 の要約結果 r3 を表示した場合には、要約結果 r3 は精度に関係するところが抽出されているが若干冗長であるため、ユーザ A は評価 2 をつけ、評価付与部 112 は事例 c3 の解として評価 2 を設定する。

## 【0050】

同様に、ユーザ B の場合を想定する。ユーザ B は、図 4 (A) に示す要約結果 r1 に対して手法に関係するところが抽出されていないために評価 3 をつけ、図 4 (B) に示す要約結果 r2 に対して手法に関係するところが抽出されていることから評価 1 をつけ、図 4 (C) に示す要約結果 r3 に対して手法に関係するところが抽出されているが若干冗長であるため評価 2 をつける。

10

## 【0051】

また、同様に、ユーザ C の場合を想定する。ユーザ C は、図 4 (A) に示す要約結果 r1 に対して精度に関係するところが抽出されているが手法に関係するところが抽出されていないため評価 2 をつけ、図 4 (B) に示す要約結果 r2 に対して手法に関係するところが抽出されているが精度に関係するところが抽出されていないため評価 2 をつけ、図 4 (C) に示す要約結果 r3 について手法および精度のいずれにも関係するところが抽出されているが若干冗長であるため評価 1 をつける。

## 【0052】

評価付与部 112 は、ユーザ B およびユーザ C ごとに要約結果 r1 ~ r3 に対する入力評価を、それぞれの事例 c1 ~ c3 の解 (評価) として設定する。

20

## 【0053】

そして、評価カスタマイズ手段 110 は、テキスト・要約 4 で与えられたテキストとその要約結果と解とを事例として解データ記憶部 120 に記憶する (ステップ S3)。

## 【0054】

図 5 に、機械学習処理および自動要約処理の流れを示す。

## 【0055】

解 - 素性対抽出部 121 は、解データ記憶部 120 から、事例ごとに解と素性の集合との組を抽出する (ステップ S11)。

## 【0056】

解 - 素性対抽出部 121 は、例えば、1) 文のなめらかさを示す情報として、k - gram 形態素列のコーパスでの存在、かかりうけ文節間の意味的整合度などを、また、2) 内容をよく表しているかどうかを示す情報として、要約前のテキストにあったキーフレーズの包含率などを、また、3) 自動要約で用いられる情報として、その文の位置やリード文かどうか、TF / IDF (TF は文書中でのその語の出現回数もしくは頻度を示す値、IDF はあらかじめ持っている多数の文書群のうち、その語が出現する文書数の逆数をいう。) 、文の長さ、固有表現・接続詞・機能語などの手がかり表現の存在などを、素性として抽出する。

30

## 【0057】

次に、機械学習部 122 は、解と素性の集合との組から、どのような素性の集合のときにどのような解になりやすいかを機械学習法により学習し、学習結果を学習結果データ記憶部 123 に記憶する (ステップ S12)。

40

## 【0058】

ここでユーザ A の処理の場合に、解データ記憶部 120 に記憶される解データの「事例：問題 解」は、

事例 c1 : テキスト - 要約結果 r1 評価 1 ,

事例 c2 : テキスト - 要約結果 r2 評価 3 ,

事例 c3 : テキスト - 要約結果 r3 評価 2

となり、機械学習部 122 は、これらの解データをもとに、どのような場合に評価 1 ~ 評価 3 になるかを機械学習で学習する。例えば、事例 c1 評価 1 や事例 c3 評価 2 から、機械学習部 122 は、精度の表現、例えば「数字 + [%]」の表現が出現すると評価が

50

高くなるなどを学習する。ここで、「数字 + [ % ]」の表現は、学習に用いる素性の例である。

【 0 0 5 9 】

また、ユーザ B の処理の場合に、「事例：問題 解」は、

事例 c 1 : テキスト - 要約結果 r 1 評価 3 ,

事例 c 2 : テキスト - 要約結果 r 2 評価 1 ,

事例 c 3 : テキスト - 要約結果 r 3 評価 2

となり、機械学習部 1 2 2 は、「手がかり表現」や「用例」などの手法に相当する専門用語が出現すると評価が高くなるように学習する。

【 0 0 6 0 】

また、ユーザ C の処理の場合に、「事例：問題 解」は、

「事例 c 1 : テキスト - 要約結果 r 1 評価 2 ,

事例 c 2 : テキスト - 要約結果 r 2 評価 2 ,

事例 c 3 : テキスト - 要約結果 r 3 評価 1 」

となり、機械学習部 1 2 2 は、精度の表現または手法に相当する表現の両方が出現すると評価が高くなるように学習する。

【 0 0 6 1 】

また、要約結果として出力される文章は短いほどよいので、それぞれの処理の場合において、文章の長さが短いほど評価が高くなるように学習する。

【 0 0 6 2 】

機械学習の手法としては、例えば、シンプルベイズ法、決定リスト法、最大エントロピー法、サポートベクトルマシン法などを用いる。

【 0 0 6 3 】

シンプルベイズ法は、ベイズの定理にもとづいて各分類になる確率を推定し、その確率値が最も大きい分類を求める分類とする方法である。

【 0 0 6 4 】

決定リスト法は、素性と分類先の組とを規則とし、それらをあらかじめ定めた優先順序でリストに蓄えおき、検出する対象となる入力を与えられたときに、リストで優先順位の高いところから入力のデータと規則の素性とを比較し、素性が一致した規則の分類先をその入力の分類先とする方法である。

【 0 0 6 5 】

最大エントロピー法は、あらかじめ設定しておいた素性  $f_j$  (  $1 \leq j \leq k$  ) の集合を  $F$  とするとき、所定の条件式を満足しながらエントロピーを意味する式を最大にするときの確率分布を求め、その確率分布にしたがって求まる各分類の確率のうち、もっとも大きい確率値を持つ分類を求める分類とする方法である。

【 0 0 6 6 】

サポートベクトルマシン法は、空間を超平面で分割することにより、2つの分類からなるデータを分類する手法である。

【 0 0 6 7 】

決定リスト法および最大エントロピー法については、以下の参考文献 1 に、サポートベクトルマシン法については、以下の参考文献 2 および参考文献 3 に説明されている。

[ 参考文献 1 : 村田真樹, 内山将夫, 内元清貴, 馬青, 井佐原均, 種々の機械学習法を用いた多義解消実験, 電子情報通信学会言語理解とコミュニケーション研究会, NCL2001-2, (2001) ]

[ 参考文献 2 : Nello Cristianini and John Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, (Cambridge University Press, 2000) ]

[ 参考文献 3 : Taku Kudoh, Tinsvml: Support Vector machines, (<http://cl.aist-nara.ac.jp/taku-ku//software/TinySVM/index.html>, 2000) ]

その後、要約を求めたいテキスト 2 が入力されると (ステップ S 1 3 ) , 要約候補生成

10

20

30

40

50

部 1 2 4 は、例えば以下に示すような処理モデルを用いて、テキスト 2 から要約候補を作成する（ステップ S 1 4）。

【 0 0 6 8 】

1) 重要文選択モデル

重要文選択モデルとは、文を単位に要約し、重要と思われる文のみを選択して残すことにより要約を実現するモデルである。このモデルの場合には、あらゆる文選択の状態をすべて解の候補とするとよい。また、すべてを解の候補とすると計算速度に支障が生じる場合には、予め備えておいた選択規則を用いて、この選択規則を満足する文の選択状態のみを解の候補とする。すなわち、所定の選択規則により候補数を減少させて処理の負荷を軽減する。なお、選択規則は、人手による規則であってもよい。

10

【 0 0 6 9 】

2) 重要箇所選択モデル

重要箇所選択モデルとは、文よりも小さいものを要約の単位として、不要なものを削除することにより要約を実現するモデルである。単位を文よりも小さいものとする以外については、上記 1) 重要文選択モデルと同様である。文よりも小さいものとして、例えば文節を用いる。すなわち、文節を単位として不要な文節を消していくことにより要約を実現する。この重要箇所選択モデルの場合には、あらゆる文節の選択の状態をすべて解の候補とする。また、すべてを解の候補とすると計算速度に支障が生じる場合には、上記 1) 重要文選択モデルと同様に、予め選択規則を用意しておき、この選択規則を満足する文の選択状態のみを解の候補とする。

20

【 0 0 7 0 】

3) 変形規則を利用したモデル

変形規則を利用したモデルとは、予め用意した変形規則を利用して要約結果を生成するモデルである。変形規則は、自動処理により獲得するか、または人手で作成しておいたものを利用する。例えば、「Xして、Yした。」を「Xした。」もしくは「Yした。」に書き換えるような変形規則を作っておき、この変形規則に従って入力「Aして、Bした。」が与えられたときに「Aした。」や「Bした。」という要約候補を生成する。

【 0 0 7 1 】

4) ランダムジェネレーションを利用したモデル

ランダムジェネレーションを利用したモデルは、例えば、入力「・・・X・・・」があったときに「・・・Y・・・」を要約候補とするようなモデルである。このとき、置き換えられる X はランダムに選ばれてもよいし、予め用意しておいた置換規則によって指定してもよい。置換規則は、人手によって生成されたものや、自動獲得したものなどを用いる。また、置き換えた先の表現 Y は、ある辞書の単語もしくは文字列の集合からランダムに選ばれてもよいし、予め用意しておいた変換規則によって指定してもよい。変換規則は、置換規則と同様、人手によって生成されたものや、自動獲得したものなどを用いる。このとき、X や Y をランダムに選ばずに、変換規則にもとづいて選ぶとすると、変形規則を利用したモデルと同じようなものになる。

30

【 0 0 7 2 】

素性抽出部 1 2 5 は、解 - 素性対抽出部 1 2 1 とほぼ同様の処理によって、入力したテキスト 2 および要約候補から素性の集合を抽出し、要約候補 - 推定解対生成部 1 2 6 へ渡す（ステップ S 1 5）。

40

【 0 0 7 3 】

そして、要約候補 - 推定解対生成部 1 2 6 は、受け取った素性の集合の場合にどのような解になりやすいかを、学習結果データをもとに推定し、すなわち、複数の要約候補のそれぞれの解（評価）とその確信度を学習結果データにもとづき算出し、要約候補と推定解との対（要約候補 - 推定解対）1 2 7 を生成する（ステップ S 1 6）。

【 0 0 7 4 】

そして、要約選択部 1 2 8 は、生成された要約候補 - 推定解対 1 2 7 から、推定解の確信度の値が最もよい要約候補 - 推定解対 1 2 7 を選択し、その要約候補を要約 3 とする（

50

ステップ S 1 7 )。

【 0 0 7 5 】

図 6 に、第 1 の実施の形態における本発明の処理装置の別の構成例を示す。

【 0 0 7 6 】

本形態では、解（分類先）として多数の候補が考えられるが、分類先の種類数が多くなり過ぎて、一般の機械学習法で処理ができない場合が生じうる。このような場合に、図 6 に示す自動要約処理装置 2 0 では、機械学習部 1 3 2 は、実際の機械学習処理において正例と負例の二種類の解（分類先）のみを考える機械学習手法を用いることにより処理が可能となる。

【 0 0 7 7 】

また、図 6 に示す自動要約処理装置 2 0 では、機械学習部 1 3 2 の学習の素性に評価という情報を用いることもできる。

【 0 0 7 8 】

自動要約処理装置 2 0 は、評価カスタマイズ手段 1 1 0 と、解データ記憶部 1 3 0 と、素性 - 解対・素性 - 解候補対抽出部 1 3 1 と、機械学習部 1 3 2 と、学習結果データ記憶部 1 3 3 と、要約候補生成部 1 3 4 と、素性 - 解候補抽出部 1 3 5 と、要約候補 - 推定解対生成部 1 3 6 と、要約選択部 1 3 8 とを備える。

【 0 0 7 9 】

解データ記憶部 1 3 0 と、素性 - 解対・素性 - 解候補対抽出部 1 3 1 と、機械学習部 1 3 2 と、学習結果データ記憶部 1 3 3 とは、特許請求の範囲に示す自動要約処理装置の機械学習処理手段を実現する処理手段である。また、要約候補生成部 1 3 4 と、素性 - 解候補抽出部 1 3 5 と、要約候補 - 推定解対生成部 1 3 6 とは、特許請求の範囲に示す要約候補生成処理手段を実現する処理手段である。

【 0 0 8 0 】

評価カスタマイズ手段 1 1 0 および要約候補生成部 1 3 4 は、図 1 に示す自動要約処理装置 1 0 の評価カスタマイズ手段 1 1 0 および要約候補生成部 1 2 4 と同様の処理を行う。

【 0 0 8 1 】

素性 - 解対・素性 - 解候補対抽出部 1 3 1 は、解データ記憶部 1 3 0 に記憶されている事例ごとに、解もしくは解候補と素性の集合との組を抽出する手段である。ここでは、解の候補は解以外の解の候補を意味し、ユーザが設定した評価を解とする。また、解と素性の集合の組を正例とし、解の候補と素性の集合との組を負例とする。

【 0 0 8 2 】

機械学習部 1 3 2 は、解もしくは解の候補と素性の集合との組から、どのような解もしくは解の候補と素性の集合のときに正例である確率や負例である確率を学習し、その学習結果を学習結果データ記憶部 1 3 3 に記憶する手段である。

【 0 0 8 3 】

素性 - 解候補抽出部 1 3 5 は、素性 - 解対・素性 - 解候補対抽出部 1 3 1 と同様の処理により、入力されたテキストおよび要約候補について、解の候補と素性の集合との組を抽出する手段である。

【 0 0 8 4 】

要約候補 - 推定解対生成部 1 3 6 は、渡された解の候補と素性の集合との組の場合に正例である確率や負例である確率を求め、正例である確率が最も大きい解を推定解として、その場合の要約候補と推定解との対（要約候補 - 推定解対）1 3 7 を生成する手段である。

【 0 0 8 5 】

要約選択部 1 3 8 は、要約候補 - 推定解対 1 3 7 の要約候補を要約 3 とする手段である。

【 0 0 8 6 】

図 7 に、自動要約処理装置 2 0 の機械学習処理および自動要約処理の流れを示す。

10

20

30

40

50

## 【0087】

素性 - 解対・素性 - 解候補対抽出部 131 は、解データ記憶部 130 から、各事例ごとに解もしくは解の候補と素性の集合との組を抽出する（ステップ S21）。そして、機械学習部 132 は、解もしくは解の候補と素性の集合との組から、どのような解もしくは解の候補と素性の集合のときに、正例である確率や負例である確率を機械学習法により学習し、学習結果を学習結果データ記憶部 133 に記憶する（ステップ S22）。

## 【0088】

その後、要約を求めたいテキスト 2 が入力されると（ステップ S23）、要約候補生成部 134 は、所定の方法でテキスト 2 から要約候補を生成する（ステップ S24）。そして、素性 - 解候補抽出部 135 は、入力したテキスト 2 および要約候補から素性の集合と解の候補との組を抽出し、要約候補 - 推定解対生成部 136 へ渡す（ステップ S25）。

## 【0089】

要約候補 - 推定解対生成部 136 は、受け取った解の候補と素性の集合との組の場合に正例や負例である確率を学習結果データをもとに推定し、正例である確率が最も大きい解の候補を推定解として、要約候補 - 推定解対 137 を生成し（ステップ S26）、要約選択部 138 は、要約候補 - 推定解対 137 の要約候補を要約 3 とする（ステップ S27）。

## 【0090】

第 1 の実施の形態では、ユーザは必要なときに自動要約処理装置 1 を使用しながら、その使用の際に出力された要約結果に対して評価 1 ~ 3 をつければよい。したがって、ユーザは、操作負担を感じることなく要約結果の評価をカスタマイズすることが可能となる。

## 【0091】

〔第 2 の実施の形態〕

図 8 に、第 2 の実施の形態における本発明の処理装置の構成例を示す。図 8 に示す自動要約処理装置 30 は、図 1 に示す自動要約処理装置 10 の評価カスタマイズ手段 110 の代わりに評価カスタマイズ手段 140 を備え、また自動要約処理装置 10 を構成する評価カスタマイズ手段 110 以外の処理手段を備える。

## 【0092】

評価カスタマイズ手段 140 は、テキスト表示部 141 と、要約編集部 142 とを備える。

## 【0093】

テキスト表示部 141 は、予め用意したテキスト 5 を表示装置（図 8 に図示しない）に表示する手段である。

## 【0094】

要約編集部 142 は、テキスト表示部 141 が表示したテキスト 5 からユーザが要約として指定した部分を抽出して、または、ユーザが指定した部分内の表現を変更して要約を編集する手段である。

## 【0095】

図 9 に、第 2 の実施の形態における評価カスタマイズ処理の流れを示す。

## 【0096】

テキスト表示部 141 は、予め用意したテキスト 5 を取り込み、表示装置に表示する（ステップ S31）。表示したテキスト 5 上でユーザに要約結果として良いと思われる部分を指定させ、ユーザが指定した範囲を受け付けて抽出する（ステップ S32）。また、指定した範囲の部分が編集されたら、その編集内容を受け付け、編集後の指定範囲部分を要約結果とする（ステップ S33）。

## 【0097】

ユーザは、表示されたテキスト上をマウスなどのポインティング・デバイスによるドラッグや、カーソルキー移動による開始位置および終了位置の指定などにより要約とする範囲を指定する。テキスト表示部 141 は、指定された範囲を、反転もしくはマーキングなどの表示により、指定されなかった範囲と区別して表示する。

10

20

30

40

50

## 【0098】

図10に、表示されるテキストの例およびユーザAが指定した範囲の例を示す。ユーザAは、破線で囲む部分「小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。」を要約としてよい部分であると指定する。要約編集部142は、図10のテキストの破線の矩形で示された部分を要約結果とする。

## 【0099】

また、ユーザBは、図11に示すように、テキストの破線で囲む部分「自然言語では、動詞を省略する」ということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現（手がかり語）と用例から補完することを行なう。」を要約として良いと指定する。また、ユーザCの場合には、図12に示すように、2つの破線の矩形で囲まれた部分「自然言語では、動詞を省略する」ということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現（手がかり語）と用例から補完することを行なう。」と部分「小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。」とを要約として良いと指定する。要約編集部142は、図11および図12に示すテキストの破線の矩形で示された部分をそれぞれ要約結果とする。

## 【0100】

なお、ユーザが指定した範囲をテキストと別に表示し、指定範囲内の表現について、ユーザが任意の箇所を削除したり、または表現を変更したりして、その内容を編集できるようにしてもよい。図13に示すように、テキスト上で指定した範囲をテキストと別に表示して、指定範囲内の語句などを削除し、追加し、訂正することができるようにする。要約編集部142は、要約決定ボタンがクリック等の操作で選択されると、その選択を受け付けて、指定範囲の内容を要約結果とする。なお、キャンセルボタンが選択された場合には、指定範囲の内容をクリアする。

## 【0101】

そして、要約編集部142は、テキスト5と要約結果とを、所定の解（良い評価）とともに解データ記憶部130に記憶する（ステップS34）。さらに、評価カスタマイズ手段140は、第1の実施の形態において、自動要約処理装置20が生成した要約、自動要約処理装置20の要約候補生成部124が生成した要約候補、人手でランダムに生成した要約などのユーザが指定した要約以外の要約に対して所定の解（悪い評価）を付与した解データも解データ記憶部130へ記憶する。

## 【0102】

以降、機械学習処理および自動要約処理の流れは、図5に示す処理の流れと同様である。ここで、機械学習部122は、それぞれのユーザごとに、図3に示すテキストと、図10～図12に示す要約結果のいずれか（すなわち、ユーザ指定範囲）と、解とする事例について学習する。

## 【0103】

図14に、第2の実施の形態における本発明の処理装置の別の構成例を示す。本形態においても、解（分類先）の種類数が多くなり過ぎて、一般の機械学習法で処理ができない場合が生じる。

## 【0104】

このため、図14に示す自動要約処理装置40では、機械学習部132は、実際の機械学習処理において正例と負例の二種類の解（分類先）のみを考える機械学習手法を用いることにより処理を可能としている。

## 【0105】

自動要約処理装置40は、図6に示す自動要約処理装置20を構成する処理手段と同様の処理手段を備え、かつ、評価カスタマイズ手段110の代わりに評価カスタマイズ手段

10

20

30

40

50

140を備えるものである。

【0106】

本形態では、ユーザに要約としてよい範囲をテキスト上で指定させるため、第1の実施の形態に比べてユーザの負担は大きい。しかし、ユーザが求める要約結果により近いものを解データ(教師)とすることができるため、ユーザが所望する要約結果をより早く出力できるように学習することができる。

【0107】

〔第3の実施の形態〕

図15に、第3の実施の形態における本発明の処理装置の構成例を示す。図15に示す自動要約処理装置50は、図1に示す自動要約処理装置10の評価カスタマイズ手段110の代わりに評価カスタマイズ手段150を備え、また他の処理手段として、自動要約処理装置10を構成する処理手段と同様の処理手段を備える。 10

【0108】

評価カスタマイズ手段150は、要約表示部151と、性質情報設定部152とを備える。

【0109】

要約表示部151は、予め用意しておいたテキスト・要約4の要約結果を表示装置(図15に図示しない)に表示する手段である。

【0110】

性質情報設定部152は、要約結果の評価にかかわる複数の性質情報を生成し、性質情報ごとの評価を設定する手段である。 20

【0111】

性質情報とは、要約結果の評価を構成する種々の性質に関する情報であり、例えば、短い文を重視しているかどうかという情報(短文重視)、要約結果に数量についての表現が含まれていることを重視しているかどうかという情報(数量表現重視)、要約結果に手法についての表現が含まれていることを重視しているかどうかという情報(手法重視)、要約結果の文体を重視しているかどうかという情報(文体重視)、要約結果の読みやすさを重視しているかどうかという情報(読みやすさ重視)などである。

【0112】

評価カスタマイズ手段150は、機械学習部122において要約結果の評価にかかわる複数の性質をそれぞれ学習することができるように、ユーザが随時必要となった評価にかかわる複数の性質情報を任意に設定できるようにして、要約結果に対するユーザの評価を複数の性質情報を用いて定義する。 30

【0113】

本形態では、性質情報設定部152で設定された性質情報の数に対応して解データ記憶部120を用意し、各性質情報ごとに機械学習を行なう。したがって、解-素性対抽出部121、機械学習部122、学習結果データ記憶部123、要約候補-推定解対生成部126の各処理手段は、性質情報の数に対応して備えられる。

【0114】

図16に、評価カスタマイズ処理の流れを示す。 40

【0115】

要約表示部151は、テキスト・要約4から取り出した要約結果を表示する(ステップS41)。性質情報設定部152は、表示した要約結果に対して複数の性質情報の項目を表示し、ユーザに各項目の値や、新規項目の設定などを促し、ユーザの入力を受け付ける(ステップS42)。

【0116】

図17および図18に、性質情報設定画面の例を示す。性質情報設定画面では、複数の位置情報のそれぞれに対応してスライダーが設けられている。ユーザはそれぞれの性質情報のスライダー上でスライドボタンを右側や左側など任意の位置を定めて性質情報ごとの評価を指定できる。例えば、ユーザは、表示された要約結果に対して「短い文重視、 50



数量表現重視，手法重視，文体重視，読みやすさ重視」などの性質情報の項目ごとに，それぞれのくらの評価になるかを，スライダー上でスライドボタンを移動させて設定する。図17および図18では，スライダーの左端から右端に向かって評価が高くなるように設定されているとする。また，ユーザはスライダーの横に任意の性質情報を入力することにより，スライダーが何を意味するかについて自由に定義できる。

**【0117】**

要約表示部151が図4(A)に示す要約結果r1を表示した場合に，図17に示すように，ユーザAは，要約結果r1が短い文なので「短い文重視」のスライドボタンを右側へ，また数量に関する表現があるので「数量表現重視」のスライドボタンを右側へ，また手法にふれていないので「手法重視」のスライドボタンを左側へ，文体と読みやすさとは

10

**【0118】**

また，要約表示部151が図4(B)に示す要約結果r2を表示した場合に，ユーザAは，図18に示すように，要約結果r2がそれほど短くないので，「短い文重視」のスライドボタンを左側へ，その他の性質情報は，まあまあよいので，その他の性質情報のスライドボタンを右側へ移動させる。

**【0119】**

そして，性質情報設定部152は，入力された性質情報ごとの値をそれぞれ解とし，その解とテキストと要約結果とを事例として性質情報ごとの解データ記憶部120に記憶する(ステップS43)。

20

**【0120】**

以降，機械学習処理および自動要約処理の流れは，図5に示す処理の流れとほぼ同様である。ここで，性質情報ごとに備えられた機械学習部122は，対応する性質情報の解データ記憶部120に記憶された事例を解データ(教師データ)として使用する。機械学習部122は，それぞれの性質情報ごとに学習を行なう。例えば，性質情報「短い文重視」については，各事例の解は，事例c1(要約結果r1)では「解=最右側」，事例c2(要約結果r2)では，「解=左側」という解ができる。機械学習部122は，これらの解データを教師データとして利用して，どういうときに短い文重視で評価されるのかを学習していく。また，その他の性質情報についても同様の学習を行なう。

30

**【0121】**

本形態では，機械学習処理後，要約候補生成部124は，入力されたテキスト2から所定の方法で要約候補を生成し，素性抽出部125は，入力テキスト2および要約候補から素性の集合を抽出する。

**【0122】**

そして，各性質情報に対応する要約候補-推定解対生成部126は，受け取った素性の集合の場合にどのような解になりやすいかを学習結果データをもとに推定し，要約候補と推定解との対(要約候補-推定解対)127を生成する。例えば，要約候補-推定解対生成部126は，複数の要約候補のそれぞれの推定解とその確信度を学習結果データにもとづき算出して，それぞれの性質情報ごとの要約候補-推定解対127を生成する。

40

**【0123】**

要約選択部128は，要約結果に対する評価の性質情報をどの程度重視するかを設定したユーザ評価設定情報7を受け付けて，要約候補-推定解対127で各性質情報の評価の値を，ユーザ評価設定情報7と比較して，最も似た要約候補-推定解対，もしくはユーザ評価設定情報7に最も適した要約候補-推定解対を選択し，その要約候補-推定解対127の要約候補を要約3とする。

**【0124】**

要約選択部128は，図17に示すような性質情報設定画面を表示して，ユーザが現在必要な要約結果の性質であるユーザ評価設定情報7を，性質情報の各項目のスライダー上のスライドボタンの位置を変更して設定するように促してもよい。

50

## 【0125】

例えば、「短い文重視」、「数量表現重視」、「手法重視」のスライドボタンを最右側に移動させ、「文体重視」、「読みやすさ重視」のスライドボタンを最左側へ移動させる場合には、ユーザは、なるべく短く、また、数量表現および手法は欠かさず、しかし、文体や読みやすさは軽視するというような性質の評価に適合する要約3を要求していることを意味するユーザ評価設定情報7となる。

## 【0126】

また、要約選択部128は、要約候補-推定解対127の簡単な選択方法として、例えば以下の式を利用して、すべての解の組合せの値Total \_\_Score を求めてもよい。

## 【0127】

$$\begin{aligned} \text{Total \_Score} = & a(\text{短い文重視}) \times \text{score}(\text{短い文重視}) \\ & + a(\text{数量表現重視}) \times \text{score}(\text{数量表現重視}) \\ & + a(\text{手法重視}) \times \text{score}(\text{手法重視}) \\ & + a(\text{文体重視}) \times \text{score}(\text{文体重視}) \\ & + a(\text{読みやすさ重視}) \times \text{score}(\text{読みやすさ重視}) \end{aligned}$$

ただし、 $a(X)$ はユーザが指定した性質情報Xのスライダーのスライドボタン位置から求まる値である。スライドボタンがスライダーの右側に位置するほど大きな値を持つとしている。 $\text{score}(X)$ は学習結果データにもとづいて算出された性質情報Xの評価の値である。要約選択部128は、この組合せ値Total \_\_Score が最も大きい要約候補-推定解対127を選択し、その要約候補を要約3として出力する。

## 【0128】

本形態では、機械学習部122で用いる要約結果を表示させてユーザに評価させるという、第1の実施の形態における処理に近い処理方法を採用した。しかし、本形態では、第2の実施の形態における機械学習の手法のように、ユーザにテキストから要約結果としてよいと思われる範囲を指定させた上で、さらに、ユーザに評価にかかわる複数の性質をスライダーなどを用いて評価させて、教師信号である解データを収集するようにしてもよい。かかる処理の場合には、同一ユーザであっても処理を行なう度に所望する要約のタイプが異なるような状況にも対処することが可能となる。また、同時に複数の性質情報を学習することが可能であるため、ユーザが評価(解)を与える際の処理負担も全体として軽減することが可能となる。

## 【0129】

本形態では、図19に示すような処理手段の構成を持つ自動要約処理装置60としてもよい。図19の自動要約処理装置60は、図6に示す自動要約処理装置20を構成する処理手段と同様の処理手段を備え、かつ評価カスタマイズ手段110の代わりに評価カスタマイズ手段150を備えるものである。

## 【0130】

自動要約処理装置60は、実際の機械学習処理において、正例と負例の二種類の解(分類先)のみを考える機械学習手法を用いることにより、機械学習での過重な処理負担を回避することができる。

## 【0131】

以上、本発明をその実施の態様により説明したが、本発明はその主旨の範囲において種々の変形が可能である。例えば、第1の実施の形態ないし第3の実施の形態のいずれの形態をも組み合わせることも可能である。

## 【0132】

〔第4の実施の形態〕

図20に、第4の実施の形態における本発明の処理の構成例を示す。図20に示す自動要約処理装置70は、解データ記憶部120と、解-素性対抽出部121と、機械学習部122と、学習結果データ記憶部123と、素性抽出部125と、解推定部160と、評価カスタマイズ手段140とを備える。

## 【0133】

10

20

30

40

50

自動要約処理装置 70 の解データ記憶部 120 , 解 - 素性対抽出部 121 , 機械学習部 122 , 学習結果データ記憶部 123 , 素性抽出部 125 および評価カスタマイズ手段 140 とは, 図 8 に示す同一番号が付与された処理手段とほぼ同様の処理を行う手段である。

【0134】

解推定部 160 は, 学習結果データ記憶部 123 の学習結果データを参照して, 素性抽出部 125 から渡された素性の集合の場合に, どのような解になり易いかを推定し, その推定解 161 を要約 3 とする手段である。

【0135】

本形態では, 解データ記憶部 120 は, テキストを問題としテキストの要約結果を解とする解データを記憶し, 機械学習部 122 は, かかる解データから抽出された解 - 素性対を用いて機械学習を行う。また, 素性抽出部 125 は, 入力されたテキスト 2 の素性を抽出して, 解推定部 160 に渡す。

【0136】

図 21 に, 第 4 の実施の形態における機械学習処理および自動要約処理の流れを示す。

【0137】

解 - 素性対抽出部 121 は, 解データ記憶部 120 から, 事例ごとに解と素性の集合との組を抽出し(ステップ S51), 次に, 機械学習部 122 は, 解と素性の集合との組から, どのような素性の集合のときにどのような解になりやすいかを機械学習法により学習し, 学習結果を学習結果データ記憶部 123 に記憶する(ステップ S52)。なお, ステップ S51, S52 の処理は, 図 5 に示すステップ S11, S12 の処理と同様である。

【0138】

その後, 要約を求めたいテキスト 2 が入力されると(ステップ S53), 素性抽出部 125 は, 解 - 素性対抽出部 121 とほぼ同様の処理によって, 入力したテキスト 2 から素性の集合を抽出し, 解推定部 160 へ渡す(ステップ S54)。そして, 解推定部 160 は, 受け取った素性の集合の場合にどのような解になりやすいかを, 学習結果データをもとに推定し, その推定解 161 を要約 3 とする(ステップ S55)。

【0139】

本形態では, テキストの要約結果を解とする解データを用いて機械学習を行い, その学習結果を参照した解推定処理において要約とすべき推定解を直接求めるようにする。

【0140】

以上, 本発明をその実施の形態により説明したが, 本発明はその主旨の範囲において種々の変形が可能であることは当然である。

【0141】

また, 本発明は, コンピュータにより読み取られ実行される処理プログラムとして実施するものとして説明したが, 本発明を実現する処理プログラムは, コンピュータが読み取り可能な, 可搬媒体メモリ, 半導体メモリ, ハードディスクなどの適当な記録媒体に格納することができ, これらの記録媒体に記録して提供され, または, 通信インタフェースを介して種々の通信網を利用した送受信により提供されるものである。

【図面の簡単な説明】

【0142】

【図 1】第 1 の実施の形態における本発明の処理装置の構成例を示す図である。

【図 2】第 1 の実施の形態における評価カスタマイズ処理の流れを示す図である。

【図 3】対象となるテキストの例を示す図である。

【図 4】要約結果の例を示す図である。

【図 5】図 1 に示す処理装置における機械学習処理および自動要約処理の流れを示す図である。

【図 6】第 1 の実施の形態における本発明の処理装置の別の構成例を示す図である。

【図 7】図 6 に示す処理装置における機械学習処理および自動要約処理の流れを示す図である。

10

20

30

40

50

- 【図 8】第 2 の実施の形態における本発明の処理装置の構成例を示す図である。
- 【図 9】第 2 の実施の形態における評価カスタマイズ処理の流れを示す図である。
- 【図 10】表示されるテキストの例およびユーザ A の指定範囲の例を示す図である。
- 【図 11】表示されるテキストの例およびユーザ B の指定範囲の例を示す図である。
- 【図 12】表示されるテキストの例およびユーザ C の指定範囲の例を示す図である。
- 【図 13】ユーザが指定した範囲の表示の例を示す図である。
- 【図 14】第 2 の実施の形態における本発明の処理装置の別の構成例を示す図である。
- 【図 15】第 3 の実施の形態における本発明の処理装置の構成例を示す図である。
- 【図 16】第 3 の実施の形態における評価カスタマイズ処理の流れを示す図である。
- 【図 17】性質情報設定画面の例を示す図である。 10
- 【図 18】性質情報設定画面の例を示す図である。
- 【図 19】第 3 の実施の形態における本発明の処理装置の別の構成例を示す図である。
- 【図 20】第 4 の実施の形態における本発明の処理装置の構成例を示す図である。
- 【図 21】図 20 に示す処理装置における機械学習処理および自動要約処理の流れを示す図である。

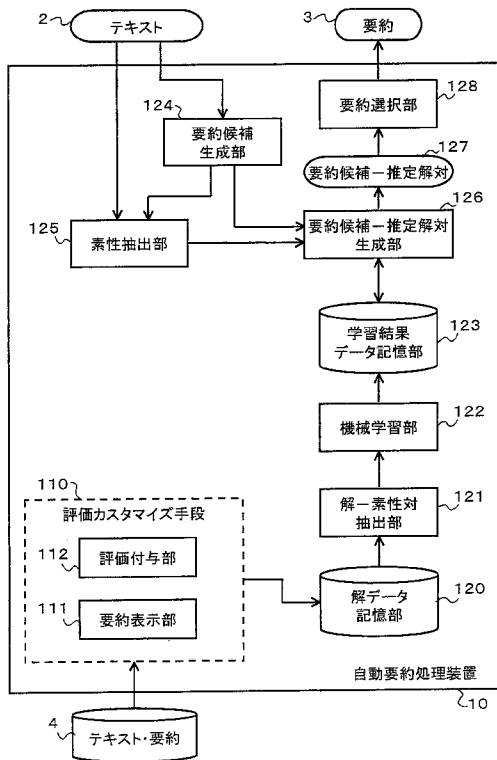
【符号の説明】

【0143】

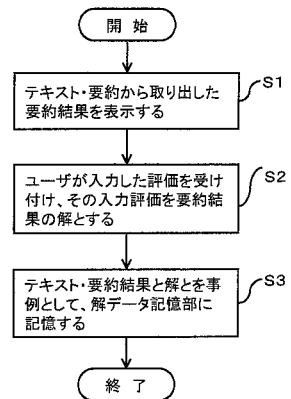
- 10 自動要約処理装置
- 110 評価カスタマイズ手段
- 111 要約表示部 20
- 112 評価付与部
- 120 解データ記憶部
- 121 解 - 素性対抽出部
- 122 機械学習部
- 123 学習結果データ記憶部
- 124 要約候補生成部
- 125 素性抽出部
- 126 要約候補 - 推定解対生成部
- 127 要約候補 - 推定解対
- 128 要約選択部 30
- 130 解データ記憶部
- 131 素性 - 解対・素性 - 解候補対抽出部
- 132 機械学習部
- 133 学習結果データ記憶部
- 134 要約候補生成部
- 135 素性 - 解候補抽出部
- 136 要約候補 - 推定解対生成部
- 137 要約候補 - 推定解対
- 138 要約選択部
- 140 評価カスタマイズ手段 40
- 141 テキスト表示部
- 142 要約編集部
- 150 評価カスタマイズ手段
- 151 要約表示部
- 152 性質情報設定部
- 160 解推定部
- 161 推定解
- 2 テキスト
- 3 要約
- 4 テキスト・要約 50

- 5 テキスト
- 7 ユーザ評価設定情報

【図1】



【図2】



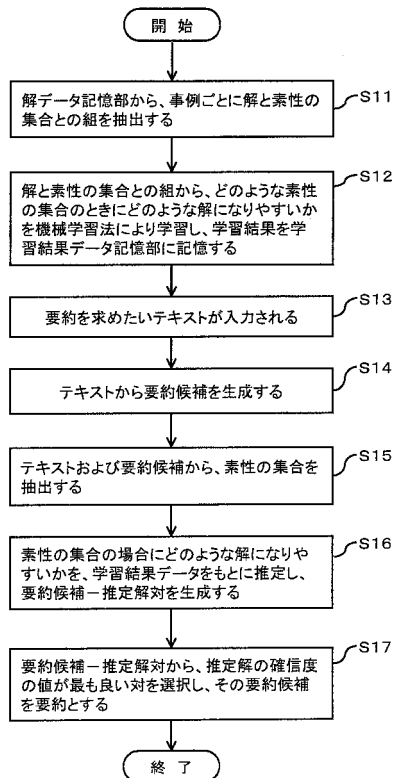
【 図 3 】

自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、コンピュータの性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。

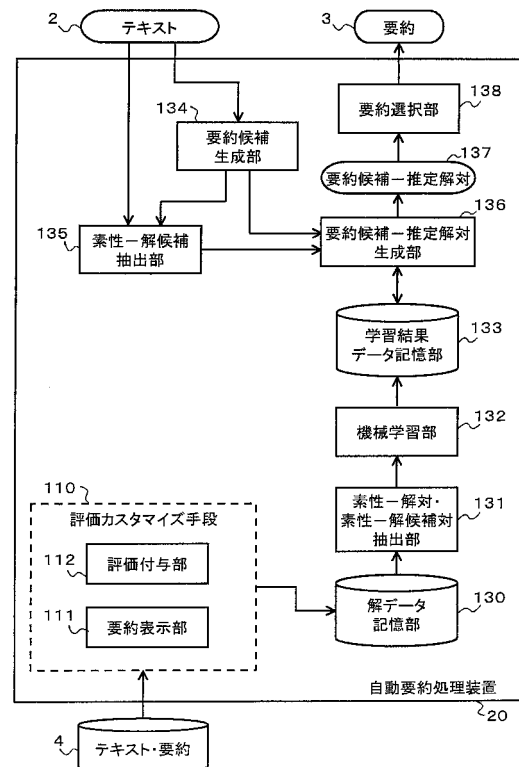
【 図 4 】

- (A) 小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。 r1
- (B) 自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。 r2
- (C) 自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。 r3

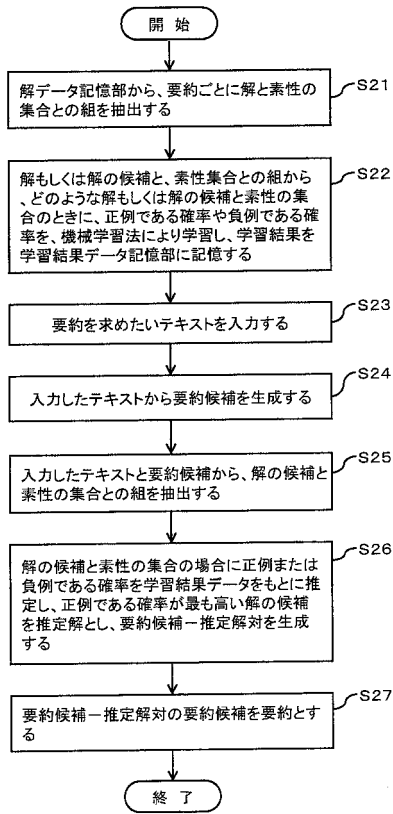
【 図 5 】



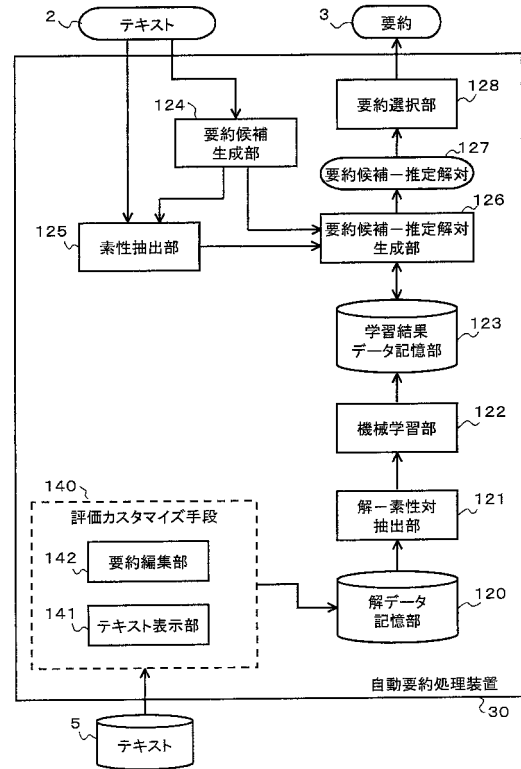
【 図 6 】



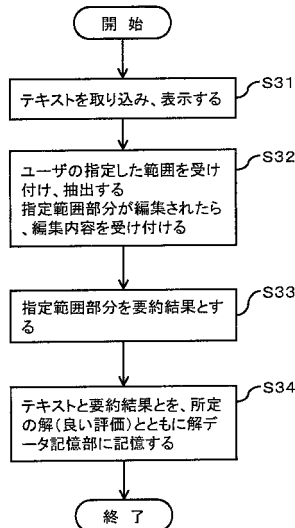
【 図 7 】



【 図 8 】



【 図 9 】



【 図 10 】

自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完を行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるかいないかなどで分類した小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。

【 図 1 1 】

自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。

【 図 1 2 】

自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。

【 図 1 3 】

テキスト:

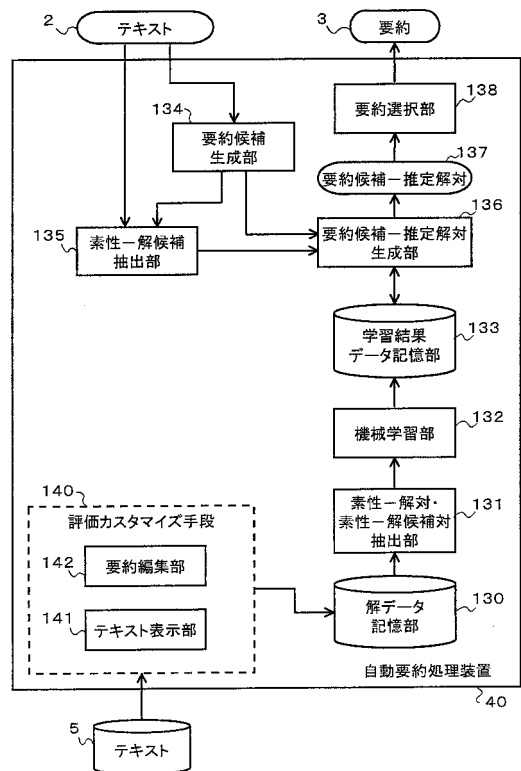
自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。

指定範囲:

自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。

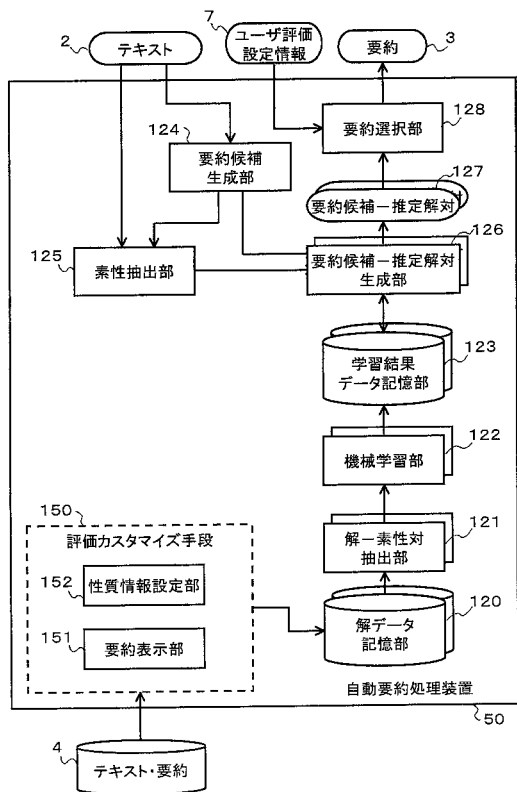
要約決定    キャンセル

【 図 1 4 】

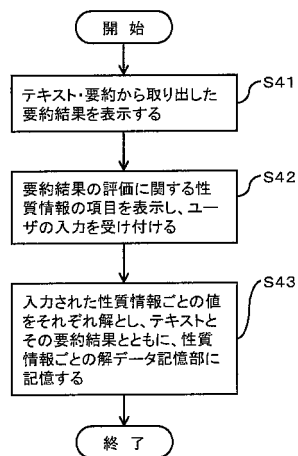




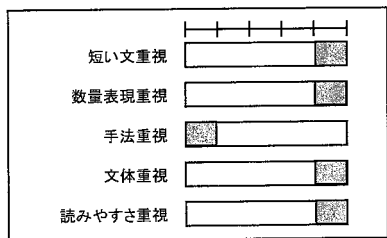
【図15】



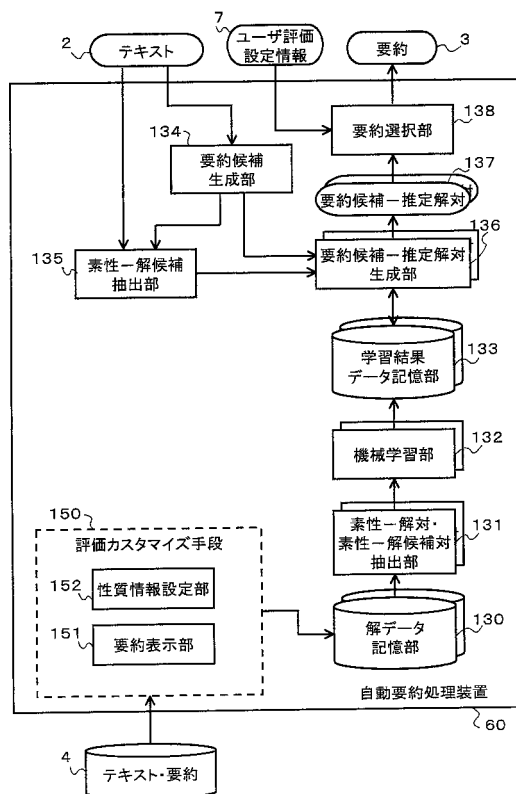
【図16】



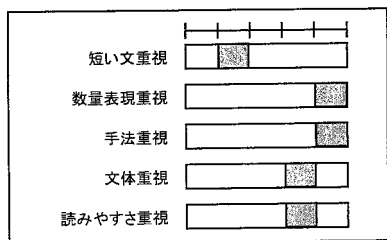
【図17】



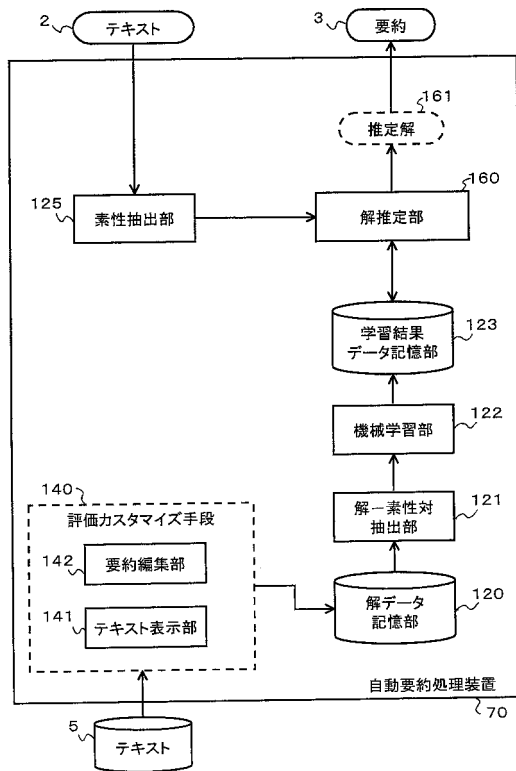
【図19】



【図18】



【図20】



【図21】

