

明 細 書

発明の名称：含意判定装置、含意判定方法及びプログラム

技術分野

[0001] 本発明は、含意判定装置、含意判定方法及びプログラムに関する。

背景技術

[0002] RTE (Recognizing Textual Entailment : 文章含意認識) と呼ばれる技術がある。これは、ある対象文章 (Text) が仮説文章 (Hypotheses) の内容を含む (含意する : Entail) か否かを判定する技術である。

[0003] 非特許文献 1 には、対象文章と仮説文章をそれぞれ、述語項構造に分割し、その構造間の距離に基づいて含意の判定を行う技術が開示されている。また、非特許文献 2 には、述語項構造への分割に代えて、全文のグラフ構造を用いて比較を行う技術が開示されている。

つまり、非特許文献 1、2 に開示された技術は、仮説文章に含まれる複数の単文のそれぞれと意味が類似する単文が、対象文章に含まれるか否かを判定することで、含意判定を行う。

先行技術文献

非特許文献

[0004] 非特許文献 1 : Wang 他、Recognizing Textual Relatedness with Predicate-Argument Structures、[Proceedings of the Conference on Empirical Methods in Natural Language Processing]、2009 年
非特許文献 2 : Haghighi 他、Robust Textual Inference via Graph Matching、[Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processi

ng」2005年

発明の概要

発明が解決しようとする課題

- [0005] しかしながら、仮説文章に含まれる複数の単文のそれぞれと対象文章に含まれる単文とが一致していたとしても、単文間の談話関係が異なる場合、対象文章が仮説文章を含意するとは言えないことがある。談話関係とは、時間列上の前後関係、因果関係、言い換えなど、単文と単文との間の役割的关系や話題の推移を示す関係である。
- [0006] 対象文章T1「Peter drank a glass of water, and afterwards went out of the room.」と、仮説文章T2「Peter left the room, and then drank something.」を比較する場合を例に説明する。対象文章T1には、単文T11「Peter drank a glass of water.」と、単文T12「Peter went out of the room.」とが含まれており、仮説文章T2には、単文T21「Peter left the room.」と単文T22「Peter drank something.」とが含まれている。
- 単文T11と単文T22とは、類似の関係にある。単文T12と単文T21とは、類似の関係にある。この場合、非特許文献1、2に記載の技術を用いると、対象文章T1は、仮説文章T2を含意すると判定される。しかしながら、仮説文章T2の「Peter went out of (left) the room.」と対象文章T1の「Peter drank a glass of water (something).」とは前後関係が異なるため、実際は、対象文章T1は、仮説文章T2を含意しない。
- [0007] 本発明の目的の一例は、上述した課題を解決する含意判定装置、含意判定方法及びプログラムを提供することにある。

課題を解決するための手段

- [0008] 本発明の実施態様に係る含意判定装置は、前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成する生成部と、前記第1および第2の情報に基づ

いて、前記第 1 の単文と前記第 2 の単文との事象の発生順の関係および前記第 3 の単文と前記第 4 の単文との事象の発生順の関係を比較することにより算出結果を得る算出部と、前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する判定部とを備える。

上記の含意判定装置は、前記第 3 の単文として、前記第 1 の単文と意味が類似する単文を、前記対象文章から抽出する抽出部をさらに備えていてもよい。

[0009] 上記の含意判定装置は、前記仮説文章を分割することによりそれぞれが述部を 1 つ有する、前記第 1 および第 2 を含む複数の単文を得て、前記対象文章を分割することによりそれぞれが述部を 1 つ有する、前記第 3 および第 4 の単文を含む複数の単文を得る分割部をさらに備えていてもよい。

[0010] 上記の含意判定装置は、第 1 の接続語と、前記第 1 の接続語によって接続される単文間の事象の発生順とを対応付けて記憶する記憶部をさらに備えていてもよい。前記生成部は、前記第 1 の接続語と前記第 1 および第 2 の単文を接続する第 2 の接続語とが同じであるか判定し、前記第 1 および第 2 の接続語が同じであると判定した場合、前記第 1 の情報における前記第 1 および第 2 の単文の順番を前記第 1 の接続語に対応付けられた事象の発生順に合せてもよい。

上記の含意判定装置は、第 1 の接続語と、前記第 1 の接続語によって接続される単文間の事象の発生順とを対応付けて記憶する記憶部をさらに備えていてもよい。前記生成部は、前記第 1 の接続語と前記第 3 および第 4 の単文を接続する第 2 の接続語とが同じであるか判定し、前記第 1 および第 2 の接続語が同じであると判定した場合、前記第 2 の情報における前記第 3 および第 4 の単文の順番を前記第 1 の接続語に対応付けられた事象の発生順に合わせてもよい。

[0011] 上記の含意判定装置において、前記抽出部は、前記第 1 の単文と前記第 3 の単文との間の第 1 の単文距離、および前記第 2 の単文と前記第 4 の単文との間の第 2 の単文距離を特定してもよい。前記判定部は、前記算出結果と前

記第 1 および第 2 の単文距離とに少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定してもよい。

[0012] 上記の含意判定装置において、前記第 1 および第 3 の単文の意味の第 1 の類似度、および前記第 2 および第 4 の単文の意味の第 2 の類似度とを算出する単文距離算出部をさらに備えていてもよい。前記算出部は、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後であることを判定してもよい。前記判定部は、前記第 1 および第 2 の類似度が高く、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後である場合、前記対象文章が前記仮説文章を含意していないと判定してもよい。

[0013] 上記の含意判定装置は、前記第 1 および第 3 の単文の意味の第 1 の類似度、および前記第 2 および第 4 の単文の意味の第 2 の類似度とを算出する単文距離算出部をさらに備えていてもよい。前記算出部は、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後であることを判定してもよい。前記判定部は、前記第 1 および第 2 の類似度が高く、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後である場合、前記対象文章が前記仮説文章を含意していないと判定してもよい。

[0014] 本発明の実施態様に係る含意判定方法は、前記仮説文章に含まれる第 1 および第 2 の単文の事象の発生順を少なくとも示す第 1 の情報を生成し、対象文章に含まれ且つ前記第 1 の単文と対応する第 3 の単文および前記対象文章に含まれ且つ前記第 2 の単文と対応する第 4 の単文の事象の発生順を少なくとも示す第 2 の情報を生成し、前記第 1 および第 2 の情報に基づいて、前記第 1 の単文と前記第 2 の単文との事象の発生順の関係および前記第 3 の単文と前記第 4 の単文との事象の発生順の関係を比較することにより算出結果を得て、前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を

含意しているか否かを判定することを含む。

[0015] 本発明の実施態様に係るプログラムは、前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成し、前記第1および第2の情報に基づいて、前記第1の単文と前記第2の単文との事象の発生順の関係および前記第3の単文と前記第4の単文との事象の発生順の関係を比較することにより算出結果を得て、

前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定することをコンピュータに実行させる。

本発明の実施態様に係る非一時的記録媒体は、上記のプログラムを記憶する。

発明の効果

[0016] 本発明の実施形態によれば、含意判定装置は、対象文章の単文と仮説文章の単文との間の談話関係が類似しない場合に、対象文章が仮説文章を含意しないと判定することができる。

図面の簡単な説明

[0017] [図1]本発明の第1の実施形態に係る含意判定装置の構成を示す概略ブロック図である。

[図2]本発明の第1の実施形態に係る含意判定装置の動作を示すフローチャートである。

[図3]本発明の第1の実施形態に係る談話関係距離の算出方法の第1の具体例を示す図である。

[図4]本発明の第1の実施形態に係る談話関係距離の算出方法の第2の具体例を示す図である。

[図5]本発明の第1の実施形態に係る談話関係距離の算出方法の第3の具体例を示す図である。

[図6]本発明の第1の実施形態に係る談話関係距離の算出方法の第4の具体例を示す図である。

[図7]本発明の第2の実施形態に係る談話関係距離の算出方法の具体例を示す図である。

[図8]本発明の第3の実施形態に係る含意判定装置の構成を示す概略ブロック図である。

[図9]本発明の第4の実施形態に係る含意判定装置の構成を示す概略ブロック図である。

発明を実施するための形態

[0018] 《第1の実施形態》

以下、図面を参照しながら本発明の実施形態について詳しく説明する。

図1は、本発明の第1の実施形態に係る含意判定装置100の構成を示す概略ブロック図である。

含意判定装置100は、与えられた対象文章と仮説文章とを解析し、対象文章が仮説文章を含意するか否かを判定する装置である。なお、対象文章及び仮説文章は、述部を複数有する複文であっても良いし、単文を含む複数の文からなる文章であっても良い。第1の実施形態に係る含意判定装置100は、文章分割部101、単文距離算出部102、類似文抽出部103、接続語記憶部104、談話関係情報生成部105、談話関係距離算出部106、含意判定部107を備える。

文章分割部101は、単に分割部101と称する場合がある。単文距離算出部102は、単に第1算出部102と称する場合がある。類似文抽出部103は、単に抽出部103と称する場合がある。接続語記憶部104は、単に記憶部104と称する場合がある。談話関係情報生成部105は、単に生成部105と称する場合がある。談話関係距離算出部106は、単に第2算出部106と称する場合がある。含意判定部107は、単に判定部107と称する場合がある。

[0019] 分割部101は、述語項構造解析及び共参照解析に基づいて、仮説文章と

対象文章をそれぞれ、述部を1つ有する複数の単文ごとに分割する。このとき、分割部101は、照応解析により単文の代名詞を参照元の名詞に置換して、さらに単文に代名詞が無い場合に適切な名詞を追加して、複文から単文への分割を行う。ここで、代名詞が無い場合の例について説明する。例えば、「太郎はリンゴを食べた後、ゴミ箱に捨てた。」という文を単純に分割すると、単文「太郎はリンゴを食べた。」と単文「太郎はゴミ箱に捨てた。」とが生成される。ここで、日本語においては、単文「太郎はゴミ箱に捨てた。」のように、日本語においては、目的語（「捨てる」の目的語）が省略されることがある。このような場合に、分割部101は、照応解析により単文「太郎は、ゴミ箱に捨てた。」に目的語「リンゴを」を追加し、単文「太郎は、リンゴをゴミ箱に捨てた。」を生成する。

[0020] 第1算出部102は、分割部101が分割した仮説文章および対象文章の単文それぞれについて、仮説文章の単文と対象文章の単文との間の意味の距離（コスト）を示す単文距離を算出する。本実施形態では、単文距離は、0以上1以下の値をとり、0に近いほど類似することを示す。第1算出部102は、単語の重複率や同義語辞書情報に基づいて単文距離を算出する。例えば、第1算出部102は、仮説文章の単文と対象文章の単文に関して、全ての単語が一致する場合に単文距離が0であると決定してもよい。第1算出部102は、同義語または類義語を有する場合に単文距離で0.5であると決定してもよい。第1算出部102は、それ以外の場合に単文距離を1とであると決定してもよい。

[0021] 抽出部103は、第1算出部102が算出した単文距離に基づいて、仮説文章から分割された単文ごとに、その単文に意味が類似する単文を、対象文章から分割された単文の中から抽出する。抽出部103は、例えば、Hungarian Algorithmを用いて、単文距離が最小となる単文の組み合わせを抽出することで、仮説文章の単文に対応する対象文章の単文を抽出することができる。また、抽出部103は、第1算出部102による算出結果に基づいて、抽出した単文ごとに、抽出した単文とその単文に対応する仮説文章の単文との

単文距離を特定する。

[0022] 記憶部 104 は、単文間を接続する接続語と、その接続語が表す談話関係とを関連付けて記憶する。例えば、記憶部 104 は、接続語「after」と、「after」が表す談話関係とを関連付けて記憶する。すなわち、記憶部 104 は、接続語「after」と、「『after』の直前に出現する単文が示す事象の前に、『after』の直後に出現する単文が示す事象が発生すること」とを関連付けて記憶する。

[0023] 生成部 105 は、仮説文章の各単文、及び抽出部 103 が抽出した各単文の間の談話関係を示す談話関係情報を生成する。具体的には、生成部 105 は、単文を文書中の出現順に並べた配列を生成し、単文間を接続する接続語に関連付けて記憶部 104 が記憶する談話関係に基づいてその配列を並べ替えることで、談話関係情報を生成する。例えば、「A after B」という文章の談話関係情報を生成する場合、生成部 105 は、まず「A」、「B」の順に単文を並べた配列を生成する。単文 A、B 間の接続語「after」が「A」が示す事象の前に「B」が示す事象が発生することを示す。このため、生成部 105 は、配列の「A」と「B」を入れ替えることで、談話関係情報を生成する。

[0024] 因果関係を示す接続語についても同様である。例えば、接続語「because」は、「『because』の直後に出現する単文が示す事象を原因として、『because』の直前に出現する単文が示す事象が発生すること」という談話関係を示す。したがって、「A because B」という文章の談話関係情報を生成する場合、生成部 105 は、「A」、「B」の順に単文を並べた配列の「A」と「B」を入れ替えることで、談話関係情報を生成する。

単文 A、B 間の接続語が「in other words」など、言い換え関係を示す場合、生成部 105 は、単文 A、B を同一のインデックスに格納することが好ましい。

[0025] 第 2 算出部 106 は、生成部 105 が生成した談話関係情報に基づいて、仮説文章の各単文、及び抽出部 103 が抽出した各単文の間の談話関係の距

離である談話関係距離を算出する。具体的には、第2算出部106は、生成部105が生成した仮説文章の各単文の談話関係情報と、抽出部103が抽出した各単文の談話関係情報とを参照し、対応する単文の位置の交差箇所の数、談話関係距離として算出する。

[0026] 判定部107は、抽出部103が特定した各単文の単文距離と、第2算出部106が算出した談話関係距離とに基づいて、対象文章が仮説文章を含意するか否かを判定する。判定部107は、例えば、単文距離と談話関係距離の総和が所定の閾値未満である場合に、対象文章が仮説文章を含意すると判定することができる。

[0027] 次に、本発明の第1の実施形態に係る含意判定装置100の動作について説明する。

図2は、本発明の第1の実施形態に係る含意判定装置100の動作を示すフローチャートである。

含意判定装置100が仮説文章及び対象文章の入力を受け付けると、分割部101は、受け付けた仮説文章及び対象文章を単文に分割する（ステップS1）。次に、第1算出部102は、仮説文章の単文と対象文章の単文の全ての組み合わせについて、単文距離を算出する（ステップS2）。

[0028] 抽出部103は、仮説文章の各単文について、仮説文章の単文に対応する対象文章の単文を抽出する（ステップS3）。このとき、抽出部103は、仮説文章の単文と抽出した対象文章の単文の間の単文距離の総和が最小になるように、対象文章から単文を抽出する。次に、抽出部103は、仮説文章の単文と抽出した単文との全ての組み合わせについて第1算出部102が算出した単文距離を特定する（ステップS4）。

[0029] 生成部105は、仮説文章の単文を、仮説文章における出現順に並べた配列を生成する。また、生成部105は、対象文章の単文を、対象文章における出現順に並べた配列を生成する（ステップS5）。次に、生成部105は、仮説文章の単文の間を接続する接続語に関連付けられた談話関係及び対象文章の単文の間を接続する接続語に関連付けられた談話関係を、記憶部10

4 から取得する（ステップ S 6）。

[0030] 生成部 105 は、取得した談話関係に基づいて仮説文章の単文を格納した配列を並べ替えて、仮説文章の談話関係情報を生成する。また、生成部 105 は、取得した談話関係に基づいて対象文章の単文を格納した配列を並べ替えて、対象文章の談話関係情報を生成する（ステップ S 7）。

[0031] 第 2 算出部 106 は、仮説文章の談話関係情報の先頭から単文を順に選択し、以下に示すステップ S 9～ステップ S 11 を実行する（ステップ S 8）。

まず、第 2 算出部 106 は、抽出部 103 が抽出した単文の中から、選択した単文に対応する単文を特定する（ステップ S 9）。次に、第 2 算出部 106 は、対象文章の談話関係情報においてステップ S 9 で特定した単文より後に格納されている単文の中に、仮説文章の談話関係情報においてステップ S 8 で選択した単文より前に格納されている単文に対応する単文が存在するか否かを判定する。つまり、第 2 算出部 106 は、ステップ S 8 で選択した単文について、仮説文章と対象文章との間に談話関係が交差する交差箇所があるか否かを判定する（ステップ S 10）。

[0032] 第 2 算出部 106 は、ステップ S 8 で選択した単文について、仮説文章と対象文章との間に談話関係が交差する交差箇所があると判定した場合（ステップ S 10：YES）、談話関係距離に 1 を加算する（ステップ S 11）。なお、談話関係距離の初期値は 0 である。他方、第 2 算出部 106 は、ステップ S 8 で選択した単文について仮説文章と対象文章との間に談話関係が交差する交差箇所がないと判定した場合（ステップ S 10：NO）、談話関係距離の値を変化させない。

[0033] 上記ステップ S 8～ステップ S 11 を、仮説文章の談話関係情報の全ての単文について実行すると、判定部 107 は、抽出部 103 が特定した単文距離の総和と第 2 算出部 106 が算出した談話関係距離とを加算した値が、所定の閾値（例えば 1）を超えるか否かを判定する（ステップ S 12）。判定部 107 は、単文距離の総和と談話関係距離とを加算した値が所定の閾値を

超えると判定した場合（ステップS 1 2 : Y E S）、対象文章が仮説文章を含意しないと判定する（ステップS 1 3）。他方、判定部 1 0 7 は、単文距離の総和と談話関係距離とを加算した値が所定の閾値を超えないと判定した場合（ステップS 1 2 : N O）、対象文章が仮説文章を含意すると判定する（ステップS 1 4）。

[0034] このように、第 1 の実施形態に係る含意判定装置 1 0 0 は、談話関係距離に基づいて対象文章が仮説文章を含意しているか否かを判定する。これにより、含意判定装置 1 0 0 は、対象文章の単文と仮説文章の単文との間の談話関係が類似しない場合に、対象文章が仮説文章を含意しないと判定することができる。

[0035] また、第 1 の実施形態に係る含意判定装置 1 0 0 は、記憶部 1 0 4 から単文の間の接続語に関連付けられた談話関係を取得し、その談話関係に基づいて単文の出現順序を入れ替えることで談話関係情報を生成する。これにより、含意判定装置 1 0 0 は、単文の出現順序と事象の順序とが異なる場合にも適切に含意判定を行うことができる。

[0036] 具体例として、対象文章 T 1 「When the sun shines, then the grass dries.」と、仮説文章 T 2 「The grass dries because of the sun shining.」を比較する場合について説明する。対象文章 T 1 には、単文 T 1 1 「The sun shines.」と、単文 T 1 2 「The grass dries.」とが含まれている。仮説文章 T 2 には、単文 T 2 1 「The grass dries.」と単文 T 2 2 「The sun shines.」とが含まれている。ここで、記憶部 1 0 4 は、接続語「then」と、「『then』の直前に出現する単文が示す事象の後に、『then』の直後に出現する単文が示す事象が発生すること」という談話関係とを関連付けて記憶する。また、記憶部 1 0 4 は、接続語「because」と、「『because』の直後に出現する単文が示す事象を原因として、『because』の直前に出現する単文が示す事象が発生すること」という談話関係を関連付けて記憶する。これにより、生成部 1 0 5 は、対象文章 T 1 の談話関係情報を並べ替えず、仮説文章 T 2 の談話関係情報の単文 T 2 1 と単文 T 2 2 とを入れ替える。

第2算出部106は、対象文章T1と仮説文章T2との間に談話関係の交差箇所が無いと判定するため、判定部107は、対象文章T1が仮説文章T2を含意すると判定することができる。

[0037] 上記ステップS8～ステップS10の処理の具体例について説明する。

図3は、本発明の第1の実施形態に係る談話関係距離の算出方法の第1の具体例を示す図である。

第1の具体例において対象文章の談話関係情報は、先頭から順に単文A1、A2、A3、A4を格納する。仮説文章の談話関係情報は、先頭から順に単文B1、B2、B3を格納する。単文B1は単文A2に対応している。単文B2は単文A3に対応している。単文B3は単文A1に対応している。

[0038] この場合、まず第2算出部106は、ステップS8において単文B1を選択する。単文B1は仮説文章の談話関係情報における先頭の単文であるため、ステップS8で選択した単文より前に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B1について交差箇所が無いと判定する。

[0039] 次に、第2算出部106は、ステップS8において単文B2を選択する。対象文章の談話関係情報において単文B2に対応する単文A3より後に格納される単文は、単文A4である。仮説文章の談話関係情報において単文B2より前に格納される単文は、単文B1である。単文A4と単文B1とは対応関係に無い。このため、第2算出部106は、ステップS10において単文B2について交差箇所が無いと判定する。

[0040] 次に、第2算出部106は、ステップS8において単文B3を選択する。対象文章の談話関係情報において単文B3に対応する単文A1より後に格納される単文は、単文A2、A3、A4である。仮説文章の談話関係情報において単文B3より前に格納される単文は、単文B1、B2である。このとき、単文A2と単文B1、及び単文A3と単文B2とは対応関係にある。このため、第2算出部106は、ステップS10において単文B3について交差箇所があると判定し、談話関係距離に1を加算する。

つまり、第1の具体例における談話関係距離は、1である。

[0041] 図4は、本発明の第1の実施形態に係る談話関係距離の算出方法の第2の具体例を示す図である。

第2の具体例において対象文章の談話関係情報は、先頭から順に単文A1、A2、A3、A4を格納する。仮説文章の談話関係情報は、先頭から順に単文B1、B2、B3を格納する。単文B1は単文A2に対応している。単文B2は単文A3に対応している。単文B3は単文A4に対応している。

[0042] この場合、まず第2算出部106は、ステップS8において単文B1を選択する。単文B1は仮説文章の談話関係情報における先頭の単文であるため、単文B1より前に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B1について交差箇所が無いと判定する。

[0043] 次に、第2算出部106は、ステップS8において単文B2を選択する。対象文章の談話関係情報において単文B2に対応する単文A3より後に格納される単文は、単文A4である。仮説文章の談話関係情報において単文B2より前に格納される単文は、単文B1である。単文A4と単文B1とは対応関係に無い。このため、第2算出部106は、ステップS10において単文B2について交差箇所が無いと判定する。

[0044] 次に、第2算出部106は、ステップS8において単文B3を選択する。単文B3に対応する単文A4は対象文章の談話関係情報における末尾の単文であるため、単文A4より後に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B3について交差箇所が無いと判定する。

つまり、第2の具体例における談話関係距離は0であり、対象文章と仮説文章との談話関係が同一であることが分かる。

[0045] 図5は、本発明の第1の実施形態に係る談話関係距離の算出方法の第3の具体例を示す図である。

第3の具体例において対象文章の談話関係情報は、先頭から順に単文A1

、A 2、A 3、A 4 を格納する。仮説文章の談話関係情報は、先頭から順に単文 B 1、B 2、B 3 を格納する。単文 B 1 は単文 A 3 に対応している。単文 B 2 は単文 A 1 に対応している。単文 B 3 は単文 A 2 に対応している。

[0046] この場合、まず第2算出部106は、ステップS8において単文B1を選択する。単文B1は仮説文章の談話関係情報における先頭の単文であるため、単文B1より前に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B1について交差箇所が無いと判定する。

[0047] 次に、第2算出部106は、ステップS8において単文B2を選択する。対象文章の談話関係情報において単文B2に対応する単文A1より後に格納される単文は、単文A2、A3、A4である。仮説文章の談話関係情報において単文B2より前に格納される単文は、単文B1である。このとき、単文A3と単文B1とが対応関係にある。このため、第2算出部106は、ステップS10において単文B2について交差箇所があると判定する。

[0048] 次に、第2算出部106は、ステップS8において単文B3を選択する。対象文章の談話関係情報において単文B3に対応する単文A2より後に格納される単文は、単文A3、A4である。仮説文章の談話関係情報において単文B3より前に格納される単文は、単文B1、B2である。単文A3と単文B1とが対応関係にある。このため、第2算出部106は、ステップS10において単文B3についても交差箇所があると判定する。

つまり、第3の具体例における談話関係距離は2である。

[0049] 図6は、本発明の第1の実施形態に係る談話関係距離の算出方法の第4の具体例を示す図である。

第4の具体例において対象文章の談話関係情報は、先頭から順に単文A1及びA2、A3、A4を格納する。対象文章では、単文A1と単文A2とが言い換えの関係にある。仮説文章の談話関係情報は、先頭から順に単文B1、B2、B3を格納する。単文B1は単文A2に対応している。単文B2は単文A1に対応している。単文B3は単文A3に対応している。

[0050] この場合、まず第2算出部106は、ステップS8において単文B1を選択する。単文B1は仮説文章の談話関係情報における先頭の単文であるため、単文B1より前に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B1について交差箇所が無いと判定する。

[0051] 次に、第2算出部106は、ステップS8において単文B2を選択する。対象文章の談話関係情報において単文B2に対応する単文A1より後に格納される単文は、単文A3、A4である。つまり、単文A2は、単文A1と同じインデックスに格納されるため、単文A1より後に格納される単文には含まれない。仮説文章の談話関係情報において単文B2より前に格納される単文は、単文B1である。単文A3、A4のいずれも単文B1と対応関係にならない。このため、第2算出部106は、ステップS10において単文B2について交差箇所が無いと判定する。

[0052] 次に、第2算出部106は、ステップS8において単文B3を選択する。対象文章の談話関係情報において単文B3に対応する単文A3より後に格納される単文は、単文A4である。仮説文章の談話関係情報において単文B3より前に格納される単文は、単文B1、B2である。単文A4は、単文B1、B2のいずれとも対応関係がない。このため、第2算出部106は、ステップS10において単文B3についても交差箇所が無いと判定する。

つまり、第4の具体例における談話関係距離は0であり、対象文章と仮説文章との談話関係が同一であることが分かる。このように、本実施形態に係る含意判定装置100は、言い換え関係にある2つの単文の順序が対象文章と仮説文章の間で入れ替わっている場合にも、適切に含意判定を行うことができる。

[0053] 《第2の実施形態》

次に、本発明の第2の実施形態について詳しく説明する。

第2の実施形態に係る含意判定装置100は、第2算出部106による談話関係距離の算出方法が、第1の実施形態と異なる。

[0054] 第2の実施形態に係る第2算出部106は、上述したステップS10で判定した交差箇所について、単文距離に基づく重みを加え、重み付けがされた交差箇所の個数を、談話関係距離として算出する。ここで、単文距離に基づく重みづけの一例について説明する。仮設文章の第1の単文および第1の単文と対応する対象文章の第2単文からなる第1の組み合わせと、第1の組み合わせと交差する、仮設文章の第3の単文および第3の単文と対応する対象文章の第4の単文からなる第2の組み合わせとが存在すると仮定する。この場合、第1の組み合わせの重みづけとしては、第1の単文と第2の単文との（第1の組み合わせの）単文距離を1から減算した値と、第3の単文と第4の単文との（第2の組み合わせの）単文距離を1から減算した値とを加算した値を用いることができる。

[0055] 図7は、本発明の第2の実施形態に係る談話関係距離の算出方法の具体例を示す図である。

第2の実施形態に係る具体例において対象文章の談話関係情報は、先頭から順に単文A1、A2、A3、A4を格納する。仮説文章の談話関係情報は、先頭から順に単文B1、B2、B3を格納する。単文B1は単文A3に対応している。単文B2は単文A2に対応している。単文B3は単文A1に対応している。単文B1－A3間の単文距離は0.5である。単文B2－A2間の単文距離は0である。単文B3－A1間の単文距離は0.5である。

[0056] この場合、まず第2算出部106は、ステップS8において単文B1を選択する。単文B1は仮説文章の談話関係情報における先頭の単文であるため、単文B1より前に格納されている単文が存在しない。そのため、第2算出部106は、ステップS10において単文B1について交差箇所が無いと判定する。

[0057] 第2算出部106は、ステップS8において単文B2を選択する。対象文章の談話関係情報において単文B2に対応する単文A2より後に格納される単文は、単文A3、A4である。仮説文章の談話関係情報において単文B2より前に格納される単文は、単文B1である。単文A3と単文B1とが対応

関係にある。第2算出部106は、単文A3-B1間の単文距離0.5を1から減算して得られる0.5と、単文A2-B2間の単文距離0を1から減算して得られる1とを加算する。第2算出部106は、このような計算から、単文B2についての談話関係距離として1.5を得る。

[0058] 次に、第2算出部106は、ステップS8において単文B3を選択する。対象文章の談話関係情報において単文B3に対応する単文A1より後に格納される単文は、単文A2、A3、A4である。仮説文章の談話関係情報において単文B3より前に格納される単文は、単文B1、B2である。単文A3と単文B1が対応関係にある。また、単文A2と単文B2が対応関係にある。第2算出部106は、単文A3-B1間の単文距離0.5と、単文A2-B2間の単文距離0とのうち小さい値を選択する。つまり、第2算出部106は、単文A2-B2間の単文距離0を選択する。第2算出部106は、単文A2-B2間の単文距離0を1から減算して得られる1と、単文A1-B3間の単文距離0.5を1から減算して得られる0.5とを加算する。第2算出部106は、この演算から、単文B3についての談話関係距離として1.5を得る。

[0059] 第2算出部106は、各単文の談話関係距離を総和することで、対象文章と仮説文章との談話関係距離を算出する。

つまり、本具体例において、談話関係距離は3である。

[0060] このように、第2の実施形態によれば、第2算出部106は、単文距離が小さい単文ほど、すなわち対応関係にある対象文章および仮説文章の単文の組み合わせの意味が類似するほど、それらの単文の組み合わせと他の対象文章および仮説文章の単文の組み合わせと間に交差が生じる場合に談話関係距離を大きくする。つまり、類似する単文について交差が生じない場合には、談話関係距離が小さい値となる。これにより、含意判定装置100は、より適切に含意判定を行うことができる。

[0061] 第2の実施形態では、交差箇所の重みの算出方法として、1から単文距離を減算する方法を例に説明をしたが、これに限られない。例えば、単文距離

の逆数をとる方法など、単文距離に対して単調減少する重みを算出する方法であれば、他の方法を用いても良い。

[0062] 《第3の実施形態》

次に、本発明の第3の実施形態について詳しく説明する。

図8は、本発明の第3の実施形態に係る含意判定装置100の構成を示す概略ブロック図である。

第3の実施形態に係る含意判定装置100は、第1の実施形態の構成に加えてさらに翻訳部108を備える。翻訳部108は、対象文章と仮説文章の言語が異なる場合に、分割部101が分割した各単文を同一の言語に統一する。つまり、翻訳部108は、対象文章から分割した各単文を仮説文章の言語に翻訳するか、仮説文章から分割した各単文を対象文章の言語に翻訳する。第1算出部102は、翻訳された単文間の単文距離を算出する。

[0063] このように、第3の実施形態によれば、含意判定装置100は、対象文章と仮説文章が異なる言語で記載されている場合にも、含意判定を行うことができる。特に、分割部101が分割した単文を翻訳するため、誤訳の発生を低減することができる。

[0064] 第1算出部102は、単語の重複率等により単文距離を算出するため、翻訳文は全文の翻訳を行わなくても良い。つまり、第1算出部102は、例えば対訳辞書を用いて単文の単語ごとに言語の変換を行っても良い。

[0065] 以上、図面を参照してこの発明の実施形態について詳しく説明してきたが、具体的な構成は上述に限られることはなく、この発明の要旨を逸脱しない範囲内において様々な設計変更等を行うことが可能である。

例えば、上述した実施形態では、談話関係距離を、交差箇所の数に基づいて算出する場合について説明したが、これに限られない。談話関係が一致しない単文の数に対して単調増加するようなその他の談話関係距離の算出方法を用いても良い。一例として、対象文章の談話関係情報を入れ替えて仮説文章の談話関係情報と同じ順序にするために必要な入れ替え回数を談話関係距離としてもよい。

[0066] 上述した実施形態では、判定部 107 が単文距離の総和と談話関係距離の和を用いて含意判定を行う場合について説明したが、これに限られない。例えば、判定部 107 は、単文距離の総和と談話関係距離の積を用いて含意判定を行っても良い。判定部 107 は、単文距離の総和及び談話関係距離をそれぞれ 0 以上 1 以下の値に正規化した後に、その和または積を用いて含意判定を行っても良い。

[0067] 上述した実施形態では、含意判定装置 100 が、単文及び談話関係の尺度として、非類似の度合いを示す距離を用いて含意判定を行う場合について説明したが、これに限られない。含意判定装置 100 は、単文及び談話関係の類似度を用いて含意判定を行っても良い。非類似の度合いを示す距離は 0 に近いほど類似していることを示す値であり、類似度は 1 に近いほど類似していることを示す値であるため、距離を用いた含意判定と類似度を用いた含意判定は等価である。

[0068] 上述した実施形態では、生成部 105 が、単文間の接続語に基づいて、単文を出現順に並べた配列を入れ替えることにより、談話関係情報を生成する場合について説明したが、これに限られない。一般的な文章は、事象の発生順に単文を並べたものが多い。このため、対象文章及び仮説文章がこのような文章である場合、含意判定装置 100 は、単文を出現順に並べたそのままの配列を談話関係情報として含意判定を行っても適切に含意判定を行うことができる。

[0069] 《第 4 の実施形態》

図 9 は、本発明の第 4 の実施形態に係る含意判定装置 100 の構成を示す概略ブロック図である。

本発明の第 4 の実施形態に係る含意判定装置 100 は、抽出部 103、生成部 105、第 2 算出部 106 及び判定部 107 を備える。

[0070] 抽出部 103 は、仮説文章に含まれる複数の単文ごとに、その単文に意味が類似する単文を、複数の単文を含む対象文章の中から抽出する。生成部 105 は、仮説文章と対象文章のそれぞれについて、単文の出現順序に基づい

て単文間の談話関係を示す談話関係情報を生成する。第2算出部106は、談話関係情報に基づいて、仮説文章に含まれる単文間の談話関係と、抽出部103が抽出した単文間の談話関係との距離である談話関係距離を算出する。判定部107は、談話関係距離に基づいて、対象文章が仮説文章を含意しているか否かを判定する。

これにより、含意判定装置100は、対象文章の単文と仮説文章の単文との間の談話関係が類似しない場合に、対象文章が仮説文章を含意しないと判定することができる。

[0071] 上述の含意判定装置100は内部に、コンピュータシステムを有してもよい。上述した各処理部の動作は、プログラムの形式でコンピュータ読み取り可能な記録媒体に記憶されており、このプログラムをコンピュータが読み出して実行することによって、上記処理が行われてもよい。コンピュータ読み取り可能な記録媒体とは、磁気ディスク、光磁気ディスク、CD-ROM、DVD-ROM、半導体メモリを含んでもよい。このコンピュータプログラムを通信回線によってコンピュータに配信し、この配信を受けたコンピュータがそのプログラムを実行するようにしても良い。

[0072] 上記プログラムは、前述した機能の一部を実現するためのものであっても良い。

上記プログラムは、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

以上、実施形態を参照して本願発明を示し説明したが、本願発明は上記実施形態に限定されない。当業者であれば、請求項によって画定される本願発明の範囲を逸脱しないで、構成や詳細に様々な変更をすることができることが理解されるであろう。

[0073] この出願は、2013年2月28日に出願された日本国特願2013-039208を基礎とする優先権を主張し、その開示の全てをここに取り込む。

産業上の利用可能性

[0074] 本発明は、含意判定装置、含意判定方法及びプログラムに適用し得る。

符号の説明

- [0075] 100…含意判定装置
101…文章分割部（分割部）
102…単文距離算出部（第1算出部）
103…類似文抽出部（抽出部）
104…接続語記憶部（記憶部）
105…談話関係情報生成部（生成部）
106…談話関係距離算出部（第2算出部）
107…含意判定部（判定部）
108…翻訳部

請求の範囲

【請求項1】 前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成する生成部と、

前記第1および第2の情報に基づいて、前記第1の単文と前記第2の単文との事象の発生順の関係および前記第3の単文と前記第4の単文との事象の発生順の関係を比較することにより算出結果を得る算出部と、

前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する判定部と

を備える含意判定装置。

【請求項2】 前記第3の単文として、前記第1の単文と意味が類似する単文を、前記対象文章から抽出する抽出部をさらに備える

請求項1に記載の含意判定装置。

【請求項3】 前記仮説文章を分割することによりそれぞれが述部を1つ有する、前記第1および第2を含む複数の単文を得て、前記対象文章を分割することによりそれぞれが述部を1つ有する、前記第3および第4の単文を含む複数の単文を得る分割部をさらに備える

請求項1または2に記載の含意判定装置。

【請求項4】 第1の接続語と、前記第1の接続語によって接続される単文間の事象の発生順とを対応付けて記憶する記憶部をさらに備え、

前記生成部は、前記第1の接続語と前記第1および第2の単文を接続する第2の接続語とが同じであるか判定し、前記第1および第2の接続語が同じであると判定した場合、前記第1の情報における前記第1および第2の単文の順番を前記第1の接続語に対応付けられた事象の発生順に合わせる

請求項 1 から請求項 3 の何れか 1 項に記載の含意判定装置。

[請求項5]

第 1 の接続語と、前記第 1 の接続語によって接続される単文間の事象の発生順とを対応付けて記憶する記憶部をさらに備え、

前記生成部は、前記第 1 の接続語と前記第 3 および第 4 の単文を接続する第 2 の接続語とが同じであるか判定し、前記第 1 および第 2 の接続語が同じであると判定した場合、前記第 2 の情報における前記第 3 および第 4 の単文の順番を前記第 1 の接続語に対応付けられた事象の発生順に合わせる

請求項 1 から請求項 3 の何れか 1 項に記載の含意判定装置。

[請求項6]

前記抽出部は、前記第 1 の単文と前記第 3 の単文との間の第 1 の単文距離、および前記第 2 の単文と前記第 4 の単文との間の第 2 の単文距離を特定し、

前記判定部は、前記算出結果と前記第 1 および第 2 の単文距離とに少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する

請求項 1 から請求項 5 の何れか 1 項に記載の含意判定装置。

[請求項7]

前記算出部は、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後であることを判定し、

前記判定部は、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発生順が後である場合、前記対象文章が前記仮説文章を含意していないと判定する

請求項 1 から請求項 6 の何れか 1 項に記載の含意判定装置。

[請求項8]

前記抽出部は、前記第 1 および第 3 の単文の意味の第 1 の類似度、および前記第 2 および第 4 の単文の意味の第 2 の類似度を算出し、

前記算出部は、前記第 1 の単文よりも前記第 2 の単文の事象の発生順が前であり、且つ前記第 3 の単文よりも前記第 4 の単文の事象の発

生順が後であるかを判定し、

前記判定部は、前記第1および第2の類似度が高く、前記第1の単文よりも前記第2の単文の事象の発生順が前であり、且つ前記第3の単文よりも前記第4の単文の事象の発生順が後である場合、前記対象文章が前記仮説文章を含意していないと判定する

請求項1から請求項7の何れか1項に記載の含意判定装置。

[請求項9]

前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、

対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成し、

前記第1および第2の情報に基づいて、前記第1の単文と前記第2の単文との事象の発生順の関係および前記第3の単文と前記第4の単文との事象の発生順の関係を比較することにより算出結果を得て、

前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する

ことを含む含意判定方法。

[請求項10]

前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、

対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成し、

前記第1および第2の情報に基づいて、前記第1の単文と前記第2の単文との事象の発生順の関係および前記第3の単文と前記第4の単文との事象の発生順の関係を比較することにより算出結果を得て、

前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する

ことをコンピュータに実行させるためのプログラム。

- [請求項11] 前記仮説文章に含まれる第1および第2の単文の事象の発生順を少なくとも示す第1の情報を生成し、
- 対象文章に含まれ且つ前記第1の単文と対応する第3の単文および前記対象文章に含まれ且つ前記第2の単文と対応する第4の単文の事象の発生順を少なくとも示す第2の情報を生成し、
- 前記第1および第2の情報に基づいて、前記第1の単文と前記第2の単文との事象の発生順の関係および前記第3の単文と前記第4の単文との事象の発生順の関係を比較することにより算出結果を得て、
- 前記算出結果に少なくとも基づいて、前記対象文章が前記仮説文章を含意しているか否かを判定する
- ことをコンピュータに実行させるためのプログラムを記憶した非一時的記録媒体。