

CLAIMS

1. A textual entailment recognition apparatus for determining whether a first text entails a second text, comprising:

a vector generation unit that generates, for each of the first text and the second text, a vector for each predicate-argument structure of the first text and the second text by using a word other than a word indicating a type of argument of a predicate in the predicate-argument structure;

a combination identification unit that compares the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text, and identifies combinations of the predicate-argument structures of the first text and the predicate-argument structure of the second text based on a result of the comparison; and

an entailment determination unit that obtains a feature amount for each of the identified combinations, and determines whether the first text entails the second text based on the obtained feature amounts.

2. The textual entailment recognition apparatus according to claim 1,

wherein the predicate-argument structure includes a predicate included in the first text or the second text, a word serving as an argument of the predicate, and a word indicating the type of argument, and

the vector generation unit uses the predicate and the word serving as the argument of the predicate to generate the vector.

3. The textual entailment recognition apparatus according to claim 1,

wherein the combination identification unit calculates a similarity between the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text, and identifies combinations of the predicate-argument structures of the first text and the predicate-argument structure of second text based on the calculated similarities.

4. The textual entailment recognition apparatus according to claim 3,

wherein the combination identification unit performs normalization processing during calculation of the similarity, in accordance with an amount of information of the vector.

5. The textual entailment recognition apparatus according to claim 1,

wherein the entailment determination unit obtains, for each of the combinations, one of a degree of word coverage and a degree of word matching for only the word serving as the argument between the predicate-argument structure of the first text and the predicate-argument structure of the second text, as the feature amount, based on a word other than a word indicating the type of argument of a predicate in the predicate-argument structure.

6. The textual entailment recognition apparatus according to claim 5,

wherein the entailment determination unit uses, in addition to the feature amount, a structural feature of the predicate-argument structure to determine whether the first text entails the second text.

7. The textual entailment recognition apparatus according to claim 6,

wherein the entailment determination unit performs determination with preference to one of the feature amount and the structural feature of the predicate-argument structure, in accordance with a structural similarity between the predicate-argument structure of the first text and the predicate-argument structure of the second text.

8. The textual entailment recognition apparatus according to claim 5,

wherein the entailment determination unit adds a weight to the feature amount based on data obtained through machine learning, when obtaining the feature amount.

9. A textual entailment recognition method for determining whether a first text entails a second text, comprising the steps of:

(a) generating, for each of the first text and the second text, a vector for each predicate-argument structure of the first text and the second text by using a word other than a word indicating the type of argument of a predicate in the predicate-argument structure;

(b) comparing the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text, and identifying combinations of the predicate-argument structures of the first text and the predicate-argument structure of the second text based on a result of the comparison; and

(c) obtaining a feature amount for each of the identified combinations, and

determining whether the first text entails the second text based on the obtained feature amounts.

10. A computer-readable recording medium having recorded thereon a program for
5 determining whether a first text entails a second text by a computer, the program including an instruction for causing the computer to execute the steps of:

(a) generating, for each of the first text and the second text, a vector for each
predicate-argument structure of the first text and the second text by using a word other
than a word indicating the type of argument of a predicate in the predicate-argument
10 structure;

(b) comparing the vector generated for each predicate-argument structure for
the first text and the vector generated for each predicate-argument structure for the
second text, and identifying combinations of the predicate-argument structures of the
first text and the predicate-argument structure of the second text based on a result of
15 the comparison; and

(c) obtaining a feature amount for each of the identified combinations, and
determining whether the first text entails the second text based on the obtained feature
amounts.

20 11. The textual entailment recognition method according to claim 9,

wherein the predicate-argument structure includes a predicate included in the
first text or the second text, a word serving as an argument of the predicate, and a word
indicating the type of argument, and, in the step (a), the predicate and the word serving
as the argument of the predicate are used to generate the vector.

12. The textual entailment recognition method according to claim 9,

wherein, in the step (b), a similarity between the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text is calculated, and combinations of the predicate-argument structures of the first text and the predicate-argument structure of second text based on the calculated similarities is identified.

13. The textual entailment recognition method according to claim 12,

wherein, in the step (b), normalization processing is performed during calculation of the similarity, in accordance with an amount of information of the vector.

14. The textual entailment recognition method according to claim 9,

wherein, in the step (c), for each of the combinations, one of a degree of word coverage and a degree of word matching for only the word serving as the argument between the predicate-argument structure of the first text and the predicate-argument structure of the second text is obtained, as the feature amount, based on a word other than a word indicating the type of argument of a predicate in the predicate-argument structure.

15. The textual entailment recognition method according to claim 14,

wherein, in the step (c), in addition to the feature amount, a structural feature of the predicate-argument structure is used to determine whether the first text entails the second text.

16. The textual entailment recognition method according to Supplementary note

15, wherein, in the step (c), determination is performed with preference to one of the

feature amount and the structural feature of the predicate-argument structure, in accordance with a structural similarity between the predicate-argument structure of the first text and the predicate-argument structure of the second text.

5 17. The textual entailment recognition method according to claim 14,
wherein, in the step (c), a weight is added to the feature amount based on data obtained through machine learning, when obtaining the feature amount.

10 18. The computer-readable recording medium according to claim 10,
wherein the predicate-argument structure includes a predicate included in the first text or the second text, a word serving as an argument of the predicate, and a word indicating the type of argument, and, in the step (a), the predicate and the word serving as the argument of the predicate are used to generate the vector.

15 19. The computer-readable recording medium according to claim 10,
wherein, in the step (b), a similarity between the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text is calculated, and combinations of the predicate-argument structures of the first text and the predicate-argument structure of second text based on the
20 calculated similarities is identified.

20. The computer-readable recording medium according to claim 19,
wherein, in the step (b), normalization processing is performed during calculation of the similarity, in accordance with an amount of information of the vector.

25

21. The computer-readable recording medium according to claim 10,

wherein, in the step (c), for each of the combinations, one of a degree of word coverage and a degree of word matching for only the word serving as the argument between the predicate-argument structure of the first text and the predicate-argument structure of the second text is obtained, as the feature amount, based on a word other than a word
5 indicating the type of argument of a predicate in the predicate-argument structure.

22. The computer-readable recording medium according to claim 21,

wherein, in the step (c), in addition to the feature amount, a structural feature of the predicate-argument structure is used to determine whether the first text entails the
10 second text.

23. The computer-readable recording medium according to claim 22,

wherein, in the step (c), determination is performed with preference to one of the feature amount and the structural feature of the predicate-argument structure, in
15 accordance with a structural similarity between the predicate-argument structure of the first text and the predicate-argument structure of the second text.

24. The computer-readable recording medium according to claim 21,

wherein, in the step (c), a weight is added to the feature amount based on data
20 obtained through machine learning, when obtaining the feature amount.

CLAIMS

1. A textual entailment determination apparatus for determining whether a first text entails a second text, comprising:

a vector generation unit that acquires a predicate-argument structure of each of the first text and the second text, and generates, for each of the first text and the second text, a vector having a character string as a component, for each predicate-argument structure by using a word other than a word indicating a type of argument of a predicate in the predicate-argument structure;

a combination identification unit that generates vectors from each of the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text by converting the character strings that are components of the vectors into numbers, calculates similarity of the two generated vectors, and identifies combinations of the predicate-argument structures of the first text and the predicate-argument structures of the second text based on the calculated similarity; and

an entailment determination unit that obtains, for each of the identified combinations, as a feature amount, either the degree of word coverage or the degree of word matching for only words serving as the arguments between the predicate-argument structures of the first text and the predicate-argument structures of the second text based on a word other than a word indicating a type of argument of a predicate in the predicate-argument structure and determines whether the first text entails the second text based on the obtained feature amounts.

2. The textual entailment recognition apparatus according to claim 1,
wherein the predicate-argument structure includes a predicate
included in the first text or the second text, a word serving as an argument
5 of the predicate, and a word indicating the type of argument, and
the vector generation unit uses the predicate and the word serving
as the argument of the predicate to generate the vector.

3. The textual entailment recognition apparatus according to claim 1 or
10 2,
wherein the combination identification unit performs normalization
processing during calculation of the similarity, in accordance with an
amount of information of the vector.

15 4. The textual entailment recognition apparatus according to any one
of claims 1 to 3,
wherein the entailment determination unit uses, in addition to the
feature amount, a structural feature of the predicate-argument structure to
determine whether the first text entails the second text.

20 5. The textual entailment recognition apparatus according to claim 4,
wherein the entailment determination unit performs determination
with preference to one of the feature amount and the structural feature of
the predicate-argument structure, in accordance with a structural similarity
25 between the predicate-argument structure of the first text and the

predicate-argument structure of the second text.

6. The textual entailment recognition apparatus according to any one of claims 1 to 5,

5 wherein the entailment determination unit adds a weight to the feature amount based on data obtained through machine learning, when obtaining the feature amount.

7. A textual entailment determination method for determining whether
10 a first text entails a second text, comprising the steps of:

(a) by a computer, acquiring a predicate-argument structure of each of the first text and the second text and generating, for each of the first text and the second text, a vector having a character string as a component, for each predicate-argument structure by using a word other than a word
15 indicating the type of argument of a predicate in the predicate-argument structure;

(b) by the computer, generating vectors from each of the vector generated for each predicate-argument structure for the first text and the vector generated for each predicate-argument structure for the second text
20 by converting the character string that are components of the vectors into numbers, calculating similarity of the two generated vectors, and identifying combinations of the predicate-argument structures of the first text and the predicate-argument structures of the second text based on the calculated similarity; and

25 (c) by the computer, obtaining, for each of the identified

combinations, as a feature amount, either the degree of word coverage or the degree of word matching for only words serving as the arguments between the predicate-argument structures of the first text and the predicate-argument structures of the second text based on a word other
5 than a word indicating a type of argument of a predicate in the predicate-argument structure and determining whether the first text entails the second text based on the obtained feature amounts.

8. The textual entailment recognition method according to claim 7,

10 wherein the predicate-argument structure includes a predicate included in the first text or the second text, a word serving as an argument of the predicate, and a word indicating the type of argument, and, in the step (a), the predicate and the word serving as the argument of the predicate are used to generate the vector.

15 9. The textual entailment recognition method according to claim 7 or 8, wherein, in the step (b), normalization processing is performed during calculation of the similarity, in accordance with an amount of information of the vector.

20 10. The textual entailment recognition method according to any one of claims 7 to 9,

wherein, in the step (c), in addition to the feature amount, a structural feature of the predicate-argument structure is used to determine
25 whether the first text entails the second text.

11. The textual entailment recognition method according to claim 10,
wherein, in the step (c), determination is performed with preference
to one of the feature amount and the structural feature of the
5 predicate-argument structure, in accordance with a structural similarity
between the predicate-argument structure of the first text and the
predicate-argument structure of the second text.

12. The textual entailment recognition method according to any one of
10 claims 7 to 11,
wherein, in the step (c), a weight is added to the feature amount
based on data obtained through machine learning, when obtaining the
feature amount.

13. A program for determining whether a first text entails a second text
by a computer, the program for causing the computer to execute the steps of:

(a) acquiring a predicate-argument structure of each of the first text
and the second text and generating, for each of the first text and the second
text, a vector having a character string as a component, for each
20 predicate-argument structure by using a word other than a word indicating
the type of argument of a predicate in the predicate-argument structure;

(b) generating vectors from each of the vector generated for each
predicate-argument structure for the first text and the vector generated for
each predicate-argument structure for the second text by converting the
25 character string that are components of the vectors into numbers,

calculating similarity of the two generated vectors, and identifying combinations of the predicate-argument structures of the first text and the predicate-argument structures of the second text based on the calculated similarity; and

5 (c) obtaining, for each of the identified combinations, as a feature amount, either the degree of word coverage or the degree of word matching for only words serving as the arguments between the predicate-argument structures of the first text and the predicate-argument structures of the second text based on a word other than a word indicating a type of
10 argument of a predicate in the predicate-argument structure and determining whether the first text entails the second text based on the obtained feature amounts.

14. The program according to claim 13,

15 wherein the predicate-argument structure includes a predicate included in the first text or the second text, a word serving as an argument of the predicate, and a word indicating the type of argument, and, in the step (a), the predicate and the word serving as the argument of the
20 predicate are used to generate the vector.

15. The program according to claim 13 or 14,

 wherein, in the step (b), normalization processing is performed during calculation of the similarity, in accordance with an amount of information of the vector.

25

16. The program according to any one of claims 13 to 15,
wherein, in the step (c), in addition to the feature amount, a
structural feature of the predicate argument structure is used to determine
whether the first text entails the second text.

5

17. The program according to claim 16,
wherein, in the step (c), determination is performed with preference
to one of the feature amount and the structural feature of the
predicate argument structure, in accordance with a structural similarity
10 between the predicate argument structure of the first text and the
predicate argument structure of the second text.

18. The program according to any one of claims 13 to 17,
wherein, in the step (c), a weight is added to the feature amount
15 based on data obtained through machine learning, when obtaining the
feature amount.