

STS Competition

イントロ

テーマ内容

- テーマ

Semantic Textual Similarity (STS / 文間類似度)

- 内容

2つのセンテンス（短文）が与えられたときに、
センテンス同士の類似度をはかる

テーマ内容

It is a dog.

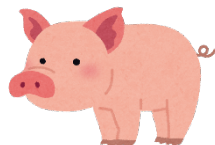
類似度: 高

It is a terrier.



類似度: 低

It is a pig.



Red alcoholic drink.

類似度: 低

Fresh orange juice.



類似度: 高

A bottle of wine.



この課題の問題点

- 新聞記事などの文書(document)単位ではなく、
文(sentence)単位での比較となる
(= 比較するために十分な情報【文脈、単語】が確保できない)



短い文章から、いかに多くの情報を得るかがカギとなる！

SemEvalについて

SemEval

- SemEvalとは...
ACL主催の競争型ワークショップ
言語処理でホットな各テーマをタスクとして設定
参加者で精度を競う

(感情分析、質問応答、構造分析etc...)

<http://alt.qcri.org/semeval2016/>

SemEval

SemEval-2016 : Semantic Evaluation Exercises International Workshop on Semantic Evaluation (SemEval-2016)

Tasks

We are pleased to announce the following exciting tasks in SemEval-2016.

Due to the high number of tasks we have organized them into tracks. The track structure is designed to show the connection between related tasks, and should make it easier for organizers to provide descriptions, datasets evaluation measures, and evaluation setups, so that it is easier for participants to participate in more than one task from the same track.

Track I. Textual Similarity and Question Answering Track

- Task 1: [Semantic Textual Similarity: A Unified Framework for Semantic Processing and Evaluation](#)
- Task 2: [Interpretable Semantic Textual Similarity](#)
- Task 3: [Community Question Answering](#)

Track II. Sentiment Analysis Track

- Task 4: [Sentiment Analysis in Twitter](#)
- Task 5: [Aspect-Based Sentiment Analysis](#)
- Task 6: [Detecting Stance in Tweets](#)

内容に応じて
複数のタスクが
用意されている

Contact Info

Organizers

Bethard, University of
Leeds, UK
Gardner, Google
Carpuat, University of
Paris, France
Jurgens, McGill Uni
Nakov, Qatar Com

Research Institute, HBKU

» Torsten Zesch, University
of Duisburg-Essen

email:

semeval-organizers@google.com

Other Info

Announcements

» June 18, 2015: Register
[here](#)

SemEval – Task1

SemEval-2016 Task 1

Task 1: Semantic Textual Similarity: A Unified Framework for Semantic Processing and Evaluation

[Register To Participate in STS 2016!](http://alt.qcri.org/semeval2016/task1/)

Semantic Textual Similarity (STS) measures the degree of equivalence in the underlying semantics of paired snippets of text. While making such an assessment is trivial for humans, constructing algorithms and computational models that mimic human level performance represents a difficult and deep natural language understanding (NLU) problem.

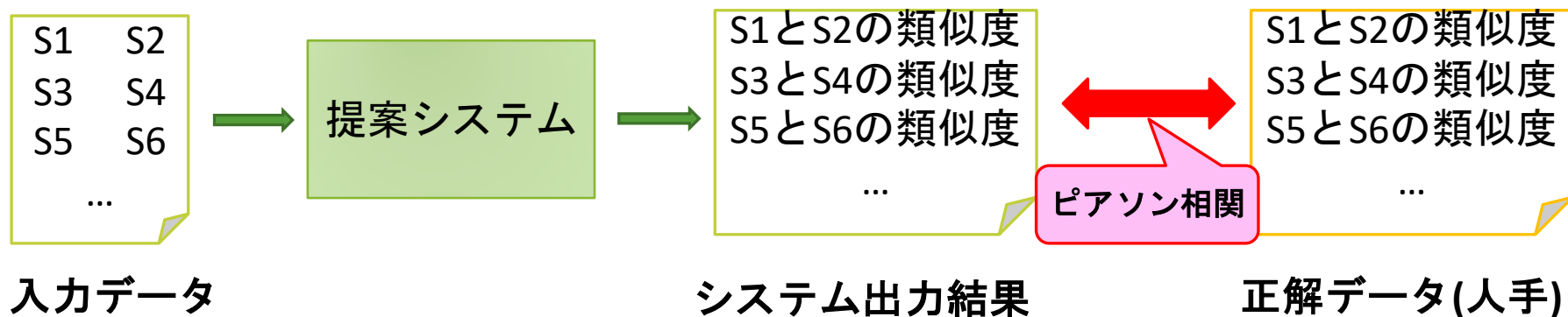
To stimulate research in this area and encourage the development of creative new approaches to modeling sentence level semantics, the STS shared task has been held annually since 2012, as part of the SemEval/*SEM family of workshops. Each year the competition brings together numerous participating teams, diverse approaches, and ongoing improvements to state-of-the-art methods.

今回はTask 1 参加を想定！

<http://alt.qcri.org/semeval2016/task1/>

Task Definition

1. 比較したい文章二つを入力
2. 文章間類似度を計算し、スコア0.0~5.0（連続値）で出力
3. 正解の点数と比較して、ピアソン相関で精度を評価する



類似度

- 正解点数(gold-standard:以下gs)は人がつけている
- 複数人で平均をとっている
- 被験者は、以下の6段階の説明を参考に点数を付与

点数	説明
5	2つのセンテンスは全く同等である
4	2つのセンテンスはほとんど同等であるが、重要でない情報が異なる
3	2つのセンテンスは大体同等であるが、いくつかの重要な情報が異なる/ 欠落している
2	2つのセンテンスは同等でない、しかし部分的に共有情報がある
1	2つのセンテンスは同等でないが同じトピックである
0	2つのセンテンスは異なるトピックである

Example

Sentence_1	Sentence_2	gs
A passenger train waiting in a station.	A passenger train sits in the station.	4.8
A group of people eat at a table outside.	A group of elderly people pose around a dining table.	2.6
Two horses standing in a field with trees in the background.	A black and white bird on a body of water with grass in the background.	0.4
A doubly decker red bus driving down the road.	A red double decker bus driving down a street.	5.0

※gs : gold-standard 正解スコア (類似度)

ゼミ課題

簡単な手順

1. 類似度計算の手法を決める
2. データ入手（使う場合は学習用・テスト用・正解スコア）
3. システム作成（使う場合は学習用データの利用）
4. 実験して類似度を算出（テストデータ利用）
5. ピアソン相関をとって精度とする（テストデータの正解スコア利用）

※ちなみにピアソン相関はSemEvalが用意したバッチファイルにかけるだけなので心配しなくて大丈夫です！

データ説明その①

データはSemEval2014のものを利用します

データ名	説明
image	Image description(PASCAL VOC2008)
OnWN	OntoNotes and Wordnet definition mappings
tweet-news	News title and tweet comments
deft-forum	Discussion forum data in the DEFT
deft-news	News article data in the DEFT project
headlines	News headlines by European Media Monitor

今回は学習用データとしてOnWN

テストデータとしてimage

を使います

データ説明その② - 学習用データ

OnWNデータについて

STS.input.OnWN.txt : 750センテンスペアがタブで区切られ格納されている

STS.gs.OnWN.txt ⇒ STS.input.OnWN.txtに対応する正解スコア750個が格納

STS.input.OnWN.txt

1行目: S₁(タブ)S₂
2行目: S₃(タブ)S₄
3行目 : S₅(タブ)S₆
...
750行目 : S₁₄₄₉(タブ)
S₁₅₀₀

STS.gs.OnWN.txt

1行目: S₁とS₂の正解類似度
2行目: S₃とS₄の正解類似度
3行目 : S₅とS₆の正解類似度
...
750行目 : S₁₄₄₉とS₁₅₀₀の正解類似度

データ説明その③ - テスト用データ

Imageデータについて

STS.input.images.txt : 750センテンスペアがタブで区切られ格納されている

STS.gs.images.txt ⇒ STS.input.images.txtに対応する正解スコア750個が格納

STS.input.images.txt

1行目: S₁(タブ)S₂
2行目: S₃(タブ)S₄
3行目 : S₅(タブ)S₆
...
750行目 : S₁₄₄₉(タブ)
S₁₅₀₀

STS.gs.images.txt

1行目: S₁とS₂の正解類似度
2行目: S₃とS₄の正解類似度
3行目 : S₅とS₆の正解類似度
...
750行目 : S₁₄₄₉とS₁₅₀₀の正解類似度

精度の測り方①

自分のシステムが出した結果(STS.output.images.txt)と
テストデータの正解データ (STS.gs.images.txt) の
ピアソン相関を測る

STS.output.images.txt

1行目: S_1 と S_2 の類似度
2行目: S_3 と S_4 の類似度
3行目 : S_5 と S_6 の類似度
...
750行目 : S_{1449} と S_{1500} の類似度

ピアソン相関

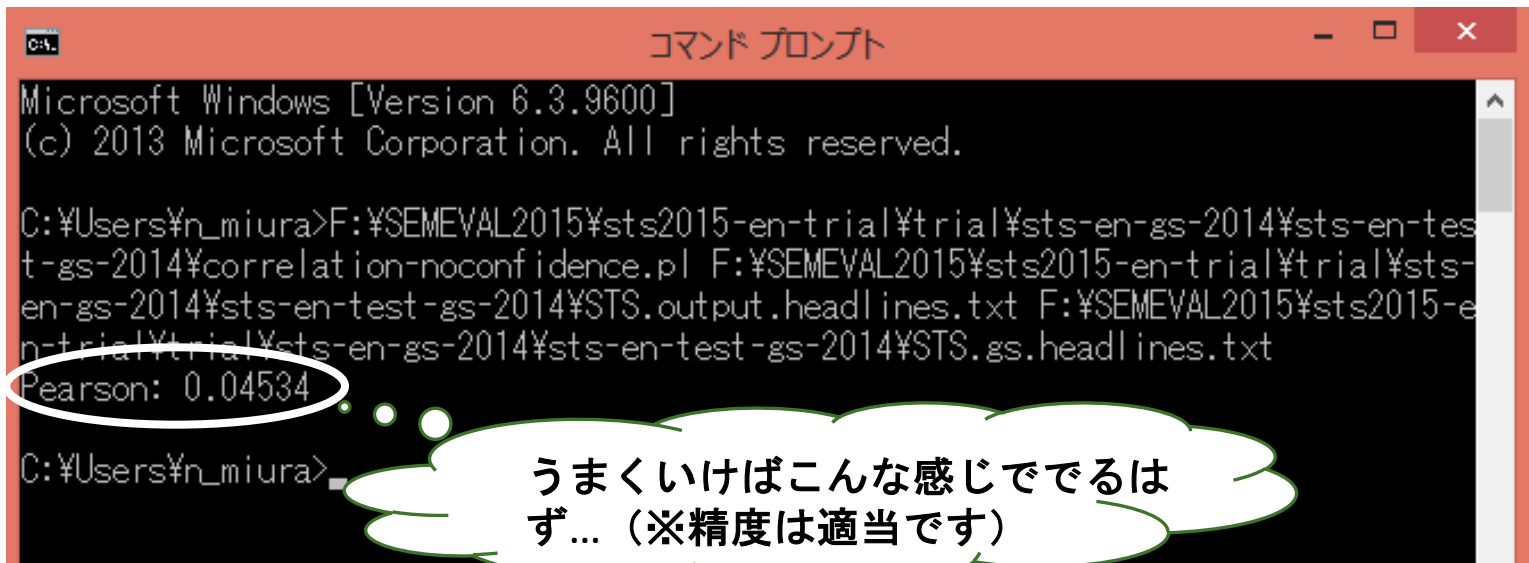
STS.gs.images.txt

1行目: S_1 と S_2 の正解類似度
2行目: S_3 と S_4 の正解類似度
3行目 : S_5 と S_6 の正解類似度
...
750行目 : S_{1449} と S_{1500} の正解類似度

精度の測り方②

コマンドプロンプト上でSemEvalが配布しているピアソン相関を計算してくれるバッチファイル(correlation-noconfidence.pl)を実行

> correlation-noconfidence.pl STS.output.images.txt STS.gs.images.txt
で実行



```
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\%n_miura>F:\%SEMEVAL2015\sts2015-en-trial\%trial\%sts-en-gs-2014\%sts-en-test-gs-2014\correlation-noconfidence.pl F:\%SEMEVAL2015\sts2015-en-trial\%trial\%sts-en-gs-2014\%sts-en-test-gs-2014\STS.output.headlines.txt F:\%SEMEVAL2015\sts2015-en-trial\%trial\%sts-en-gs-2014\%sts-en-test-gs-2014\STS.gs.headlines.txt
Pearson: 0.04534

C:\Users\%n_miura>
```

うまくいけばこんな感じででは... (※精度は適当です)