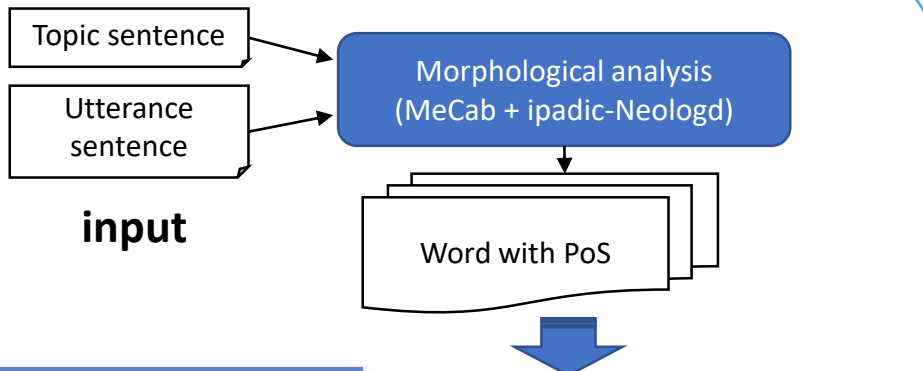


FU-01 Team's Classification of Fact-checkable Opinions in NTCIR-14 QA Lab-PoliInfo Task

Souichi Furukawa, Yuto Naritomi, **Hokuto Otake**, Toshifumi Tanabe and Kenji Yoshimura

We proposed two methods:

(i) Rule-based



Fact-checkability

Assume the input as fact-checkable if its utterance includes at least one of the keywords as described the left table.

おり	ori (and)
例	rei (example)
から	kara (because)
であり	de-ari (and)
理由	riyu (reason)

Topic: カジノを含む統合型リゾートを推進するべきである
賛成者のうち約七割が、カジノによる経済波及効果や雇用の創出を賛成の理由としています。

Relevance

Assume the input as being relevant with the topic if its utterance includes at least one of the nouns in the topic.

Example:

Topic: カジノを含む統合型リゾートを推進するべきである
・・・昨年オープンしたマリーナ・ベイ・サンズなどは、**カジノ**施設を含めたホテル、・・・

Stance

Assume a sentiment polarity score that is +1 or -1 if a recorded sentiment polarity in Japanese Sentiment Dictionary is positive or negative, respectively. The system considers the input as one of the three stances as described the left table.

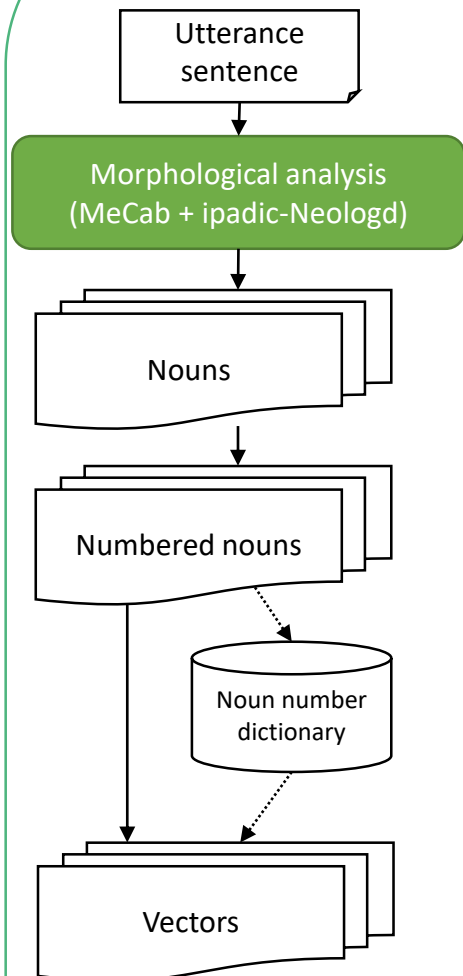
Stance	Sum of scores
approving	$\geq +1$
opposite	≤ -1.7
other	otherwise

Example:

この土壌**汚染**対策によって、
-1
豊洲新市場用地の**安全性**は**揺るぎない**ものになったと思えますが~
+1 +1

Sum of scores = -1 + 1 + 1 + ...

(ii) MaxEnt classifier based



Example: input

昨年度の知事のカジノについての発言以降、昨年6月、そして...

Extract nouns from the result of morphological analysis of the utterance sentence.

Example: nouns

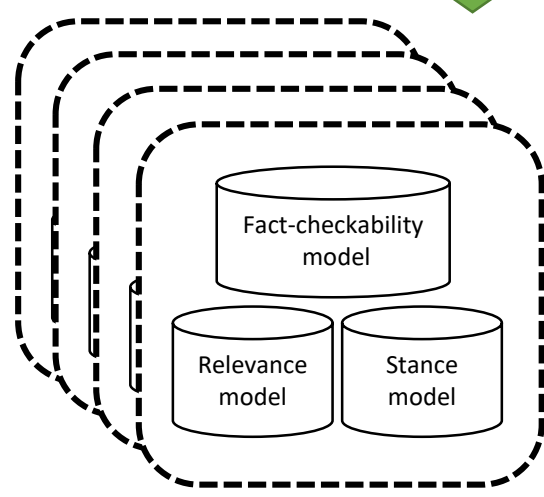
昨年度, 知事, カジノ, 発言, 以降, 昨年, 6月...

Example: numbered nouns

昨年度:1, 知事:2, カジノ:3, 発言:4, 以降:5, 昨年:6, 6月:7...

Example: sentence vector

[1, 2, 3, 4, 5, 6, 7, ...]

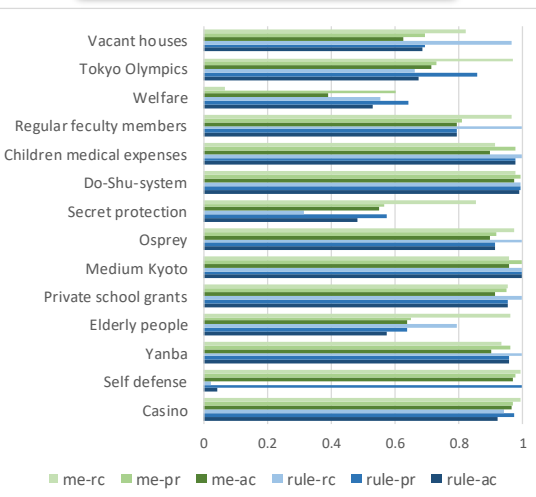


MaxEnt classification models for each topic.

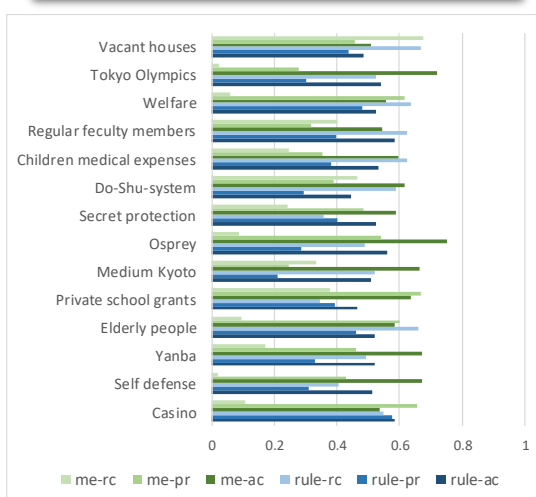
We also propose a method based on a maximum entropy classifier for estimating the three kinds of labels.

We construct models from pairs of a sentence vector and the label using "scikit-learn" library.

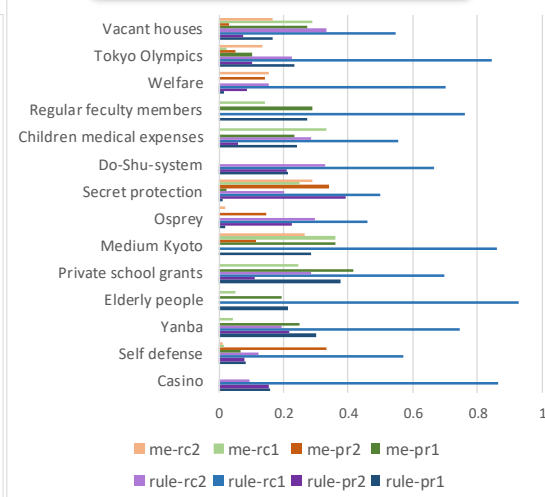
Relevance result



Fact-checkability result



Stance result



The rule-based accuracy of relevance of "Self defense" topic is very low. The reason of this is the difference between the results of the two morphological analysis, "集団的自衛" and "集団的自衛権".

As one of the reasons for lowering accuracy of ME, it is conceivable that the bias of data size between training data and test data is different.