

Appendix A: First version of disagreement guidelines

1 Introduction

The annotated SNLI corpus contains 10.000 sentence pairs, consisting of a premise (S1) and a hypothesis (S2). Annotators were asked to classify the relation between the S1 and S2 as entailment, contradiction or neutral. These are so-called inference relations. In the original guidelines, these labels are defined as:

Entailment = S2 is a definitely true description of S1

Neutral = S2 might be a true description of S1

Contradiction = S2 is a definitely false description of S1

The 10.000 sentence pairs are all classified by five annotators. Not all annotators classify the sentence pairs the same way. Sometimes, there is disagreement. Disagreement is defined as a 2vs3 or 3vs2 classifying of a sentence pair. For example, a particular pair is classified as entailment by three persons, and as neutral by two persons. In 1203 sentence pairs (12%) disagreement occurs.

This task is aimed at classifying the different disagreements. Earlier analysis on our side yielded three classes that are hypothesized to cause disagreement: suggestive information, referential ambiguity and misunderstanding. These classes are explained below.

2 Task description

You are asked to read 25 sentence pairs. Classify them into one of the three labels explained below. If you are certain neither of the three categories apply to the sentence pair, select the option “undefinable”.

Suggestive information

Information in the premise (S1) suggests that there is a relation between the premise (S1) and the hypothesis (S2). It could also be the case that the world of the annotator plays a role in their judgment.

Examples:

Premise (S1): “A man selling donuts to a customer during a world exhibition event held in the city of Angeles”

Hypothesis (S2): “A man selling donuts to a customer during a world exhibition event while people wait in line behind him.”

Explanation: The information in the premise (“during a world exhibition event”) suggests that there are multiple people present at the location where the man is selling donuts, which suggests the hypothesis, namely that there are people waiting in line.

Premise (S1): “Two black little boys hanging over an upside down bicycle, one examining the pedal.”

Hypothesis (S2): “The children are outside.”

Explanation: The information in the premise (“hanging over an upside down bicycle”) suggests that they are either outside or inside, depending on the world of the annotator, influencing their judgment on the hypothesis.

Referential ambiguity

It is not specified whether an event/object/person in the hypothesis (S2) refers to the same event/object/person in the premise (S1).

Examples:

Premise (S1): “An older gentleman finds it humorous that he is getting his picture taken while doing his laundry.”

Hypothesis (S2): “a blond man laughs while doing laundry.”

Explanation: It is unclear whether the hypothesis refers to the “older gentleman” in the premise. Differences in referencing causes different judgment by annotators.

Premise (S1): “A woman is talking on the phone while standing next to a dog.”

Hypothesis (S2): “A woman is walking her dog.”

Explanation: It is unclear whether “a woman” in the premise is the same “a woman” in the hypothesis. Differences in referencing causes different judgment by annotators.

Misunderstanding

The annotator has misread the premise (S1) and is not able to give a correct judgment about the hypothesis (S2). It could also be the case that the premise (S1) gives so many details, that the annotator might be distracted by them while judging the hypothesis (S2).

Examples:

Premise (S1): “A man in a blue shirt standing in front of a garage-like structure painted with geometric designs.”

Hypothesis (S2): “A man is wearing a blue shirt.”

Explanation: The hypothesis is a clear entailment of the premise. The annotator was not able to give a correct judgment, since a disagreement between annotators was found.

Premise (S1): “Three young boys one is holding a camera and another is holding a green toy all are wearing t-shirt and smiling.”

Hypothesis (S2): “The boys are all wearing yellow.”

Explanation: The hypothesis should be judged as neutral by the annotators, but a disagreement occurred. This can be caused by a misreading of the premise, since it contains a lot of irrelevant information, and the color green for another object.

Appendix B: Second version of disagreement guidelines

1 Introduction

The annotated SNLI corpus contains 10.000 sentence pairs, consisting of a premise (S1) and a hypothesis (S2). Annotators were asked to classify the relation between the S1 and S2 as entailment, contradiction or neutral. These are so-called inference relations. In the original guidelines, these labels are defined as:

Entailment = S2 is a definitely true description of S1

Neutral = S2 might be a true description of S1

Contradiction = S2 is a definitely false description of S1

The 10.000 sentence pairs are all classified by five annotators. Not all annotators classify the sentence pairs the same way. Sometimes, there is disagreement. Disagreement is defined as a 2vs3 or 3vs2 classifying of a sentence pair. For example, a particular pair is classified as entailment by three persons, and as neutral by two persons. In 1203 sentence pairs (12%) disagreement occurs.

2 Task description

Your task is to read the sentence pairs and classify them into one of the categories explained below.

1) When reading the sentence pairs, think about the task of the annotator: what is the relation between the two sentences (neutral, entailment, or contradiction)?

2) The sentence pairs are disagreements between annotators, so for every sentence pair, two labels are possible. Try to think of what labels the annotators chose and why, so you can find out why there is a disagreement between annotators.

3) Categorize the disagreement pairs into one of the classes below:

Categories

Misunderstanding

The annotator has misread the premise (S1) and is not able to give a correct judgment about the hypothesis (S2). It could also be the case that the premise (S1) gives so many details, that the annotator might be distracted by them while judging the hypothesis (S2).

Examples:

Premise (S1): "A man in a blue shirt standing in front of a garage-like structure painted with geometric designs."

Hypothesis (S2): "A man is wearing a blue shirt."

Explanation: The hypothesis is a clear entailment of the premise. The annotator was not able to give a correct judgment, since a disagreement between annotators was found.

Premise (S1): "Three young boys one is holding a camera and another is holding a green toy all are wearing t-shirt and smiling."

Hypothesis (S2): "The boys are all wearing yellow."

Explanation: The hypothesis should be judged as neutral by the annotators, but a disagreement occurred. This can be caused by a misreading of the premise, since it contains a lot of irrelevant information, and the color green for another object.

Suggestive information

Information in the premise (S1) suggests that there is a relation between the premise (S1) and the hypothesis (S2). It could be the case that the world of the annotator plays a role in their judgment.

Examples:

Premise (S1): "A man selling donuts to a customer during a world exhibition event held in the city of Angeles."

Hypothesis (S2): "A man selling donuts to a customer during a world exhibition event while people wait in line behind him."

Explanation: The information in the premise ('during a world exhibition event') suggests that there are multiple people present at the location where the man is selling donuts, which suggests the hypothesis, namely that there are people waiting in line. The disagreement is caused by either accepting, rejecting or overlooking this suggestion.

Premise (S1): "Two black little boys hanging over an upside down bicycle, one examining the pedal."

Hypothesis (S2): "The children are outside."

Explanation: The information in the premise ('hanging over an upside down bicycle') suggests that they are either outside or inside, depending on the world of the annotator, influencing their judgment on the hypothesis.

Referential ambiguity

It is not specified whether an event/object/person in the hypothesis (S2) refers to the same event/object/person in the premise (S1).

If S2 features the same noun as S1 does, but it is modified by an adjective (such as 'happy'), you can assume that it is about the same individual, therefore there is no referential ambiguity:

A man – A happy man = no referential ambiguity

A man – A police officer = referential ambiguity

If S2 features a synonym of a noun in S1, which intuitively refers to the S1 noun, you can assume that it is about the same individual, therefore there is no referential ambiguity:

A man – A guy = no referential ambiguity

A man – A tourist = referential ambiguity

Appendix C: Visual representation

