

WHAT'S THE ANSWER: DIALOGUE ANNOTATION

María Andrea Cruz Blandón, Gosse Minnema & Aria Nourbakhsh

JUNE 6TH, 2018



SUPERVISED PROJECT
WITH MAXIME AMBLARD & MARIA BORTICHEV

OVERVIEW

OBJECTIVES

1. Design an **annotation schema** for classifying questions and answers
2. Write an **annotation guide** and **manually annotate** dialogues in different languages
3. Explore **machine learning approaches** to automate our work

RESEARCH CONTEXT

- **SLAM Project:** computational models of the speech of patients with schizophrenia
- **Our contribution:**
 - ⇒ Preprocessing for automatic dialogue modeling systems
 - ⇒ Analyze unimpaired speech, baseline for research into questions/answers in speech of people with schizophrenia

CONCLUSIONS

General conclusions

Annotation schema and guide work well for English, need extra adaptation for other languages

Agreement scores are moderate; this is expected because the task is subtle and complex

Machine learning: not enough data; reasonable results but poor generalization

Future work

Machine learning for answer classification + answer span detection

Produce more annotated data (using semi-automatic tagging?)

Explore unsupervised machine learning approaches

ANNOTATION SCHEMA

QUESTION TYPES

Balance between **comprehensiveness** and **simplicity**

Corpus-based approach: start with basic assumptions and expand schema to fit the data

Based on **forms** (syntax) and **functions** (semantics/pragmatics)

Wh-questions and disjunctive questions get **feature** annotations that indicate the semantic role of the asked information (e.g. 'AGENT', 'THEME', 'REASON')

Tag	Name	Example
YN	Yes/no question	Did you get a new sweater?
WH	Wh-question	What are your plans now?
DQ	Disjunctive question	Is the wedding on Sunday or is it Saturday?
PQ	Phatic question	Right? / Oh yeah? / You know?
CS	Completion suggestion	A: It includes heat and uhm, I think ... B: Water?

ANSWER TYPES

Tag	Name	Example
PA, NA, FA	Positive/negative answer	Yeah. / No wait, this is a different one.
FA	Feature answer	Q: How's Gran Mary? A: She's fine.
PHA	Phatic answer	I know.
UA	Uncertainty answer	Uhm, I don't even know for sure.
UT	Unrelated topic	Q: When will you guys get off? A: My exam is like ... I don't know.
DA	Deny assumption	Q: What are your plans now? A: I don't have any plans.

OTHER TAGS

Quotations

Tag	Name	Example
QQ	Quoted question (or answer)	- Christina was like "what's so funny?" - I'm like, "I can't even tell you because ..."
NQ	Non-quoted question (or answer)	n/a

Complexity

Tag	Name	Example
SQ	Single question	Do you have a white shirt?
MQ	Multiple question	Wasn't there a musical about the Moulin Rouge, or was that something else?

ANNOTATIONS

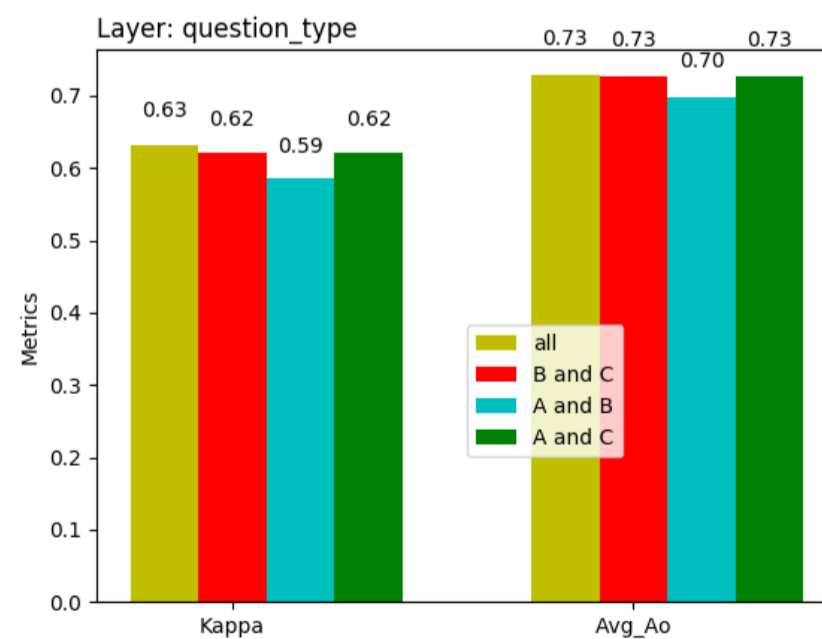
CORPORA & SOFTWARE

Annotated dialogues

- **SCoSE (Saarbrücken Corpus of Spoken English), part I**
45 000 words, conversations between US college students
Annotated 4963 lines, found 422 questions and 289 answers
- **CallFriend**
20 000 words, phone calls of speakers of Colombian Spanish
Annotated 2648 lines, found 192 questions and 122 answers
- **CGN (Corpus of Spoken Dutch)**
7000 words, phone calls of speakers from the Netherlands
Annotated 935 lines, found 87 questions and 72 answers

Software: ELAN

INTER-ANNOTATOR AGREEMENT

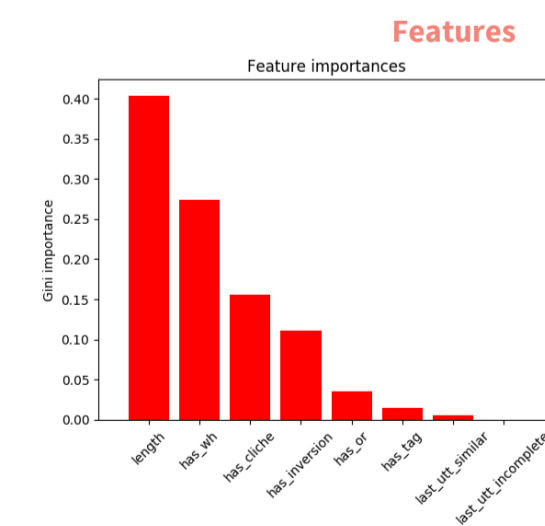


Kappa Score: assumes that every annotator has a different distribution

Observed agreement (A_o): how often did annotators choose the same category?

MACHINE LEARNING (QUESTION CLASSIFICATION)

DECISION TREE



Automatic feature extraction using simple **heuristics**

Gini coefficient: number of times feature is used to make a decision

	Accuracy	Precision	Recall	F1
Seen	0.80	0.90	0.71	0.74
Unseen	0.73	0.56	0.60	0.58

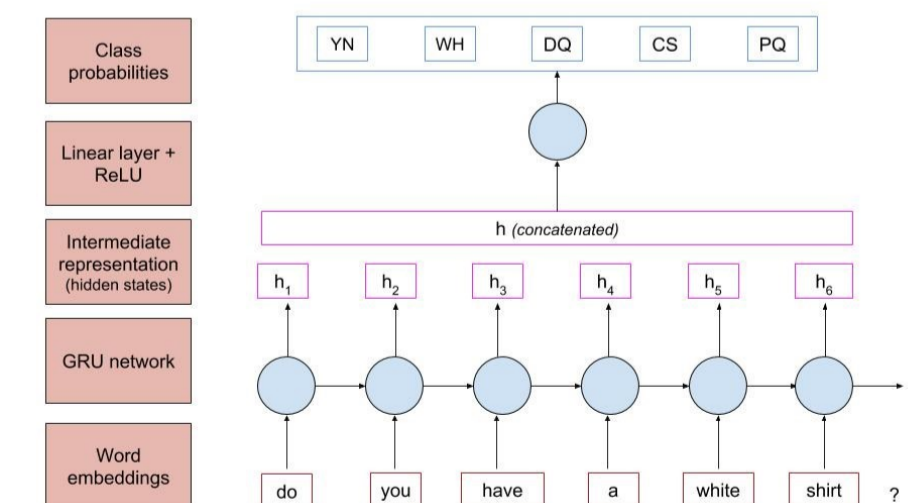
NEURAL NETWORKS

RNN Classifier

Not enough data (~250 samples), so no real results

GRU as encoder, linear classifier as decoder

Other models: bag-of-words, multi-layer perceptron



Results

	Accuracy	Precision	Recall	F1
Seen	0.92	0.76	0.70	0.72
Unseen	0.40	0.18	0.18	0.16