Predicting Multimodal Clarification

Summary & Future work

## Learning Multimodal Clarification Strategies

#### Verena Rieser<sup>1</sup> Ivana Kruijff-Korbayová<sup>1</sup> Oliver Lemon<sup>2</sup>

<sup>1</sup>Department of Computational Linguistics, Saarland University

> <sup>2</sup>School of Informatics, University of Edinburgh

In affiliation with the TALK Project http://www.talk-project.org/ CALK @

## Example of a Multimodal Clarification Strategy

- User: Show me all songs with the title "Believe".
- Wizard: I am searching for titles called "Believe".
  - User: OK.
- Wizard: There are too many hints. Do you know the name of the artist or the album ?
  - User: No. Please show me a selection.
- Wizard: I'm searching for a selection of this title.
  - User: Yes. All right.
- Wizard: Should I display all albums from 2004 containing this title?
  - User: Yes. All right.
- Wizard: I'm showing you the list.
  - User: OK. ... [clicks] Show me the song from the selected album. Play it.

ERSITÄT LANDES

Predicting Multimodal Clarification

Summary & Future work

## Why Multimodality?

Multimodality helps to reduce:

- + interpretation uncertainty (Orviatt, 2002)
- + the cognitive load of the user (Oriviatt et al., 2004)

The use of multimodality is context dependent.

- safety in the in-car domain
- high number of hints in the data base vs. short term memory
- type of interpretation uncertainty
- user model
- etc.



Predicting Multimodal Clarification

Summary & Future work

#### Outline

#### Framework Bootstrapping Reinforcement Learning from WOZ Data

#### **Predicting Multimodal Clarification**

The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Outline

#### Framework

#### Bootstrapping Reinforcement Learning from WOZ Data

#### Predicting Multimodal Clarification

The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Outline

#### Framework

#### Bootstrapping Reinforcement Learning from WOZ Data

#### Predicting Multimodal Clarification

The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



Predicting Multimodal Clarification

Summary & Future work

#### Outline

#### Framework Bootstrapping Reinforcement Learning from WOZ Data

Predicting Multimodal Clarification The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

## **Thesis Goals**

Overall goal:

We want to learn a clarification strategy which is more natural, context dependent, and flexible, while maximising user satisfaction.

Sub-goals

1. Investigate human behaviour given understanding uncertainties.

ightarrow Collect data on possible strategies in WOZ experiment.  ${}^{m {O}}$ 

2. Learn a strategy that reflects human behaviour depending on the context.

→ "Bootstrap" an initial policy using SL.

## **Thesis Goals**

Overall goal:

We want to learn a clarification strategy which is more **natural**, context dependent, and flexible, while maximising user satisfaction.

Sub-goals

1. Investigate human behaviour given understanding uncertainties.

ightarrow Collect data on possible strategies in WOZ experiment.  ${m {\it O}}$ 

2. Learn a strategy that reflects human behaviour depending on the context.

→ "Bootstrap" an initial policy using SL.

## **Thesis Goals**

Overall goal:

We want to learn a clarification strategy which is more **natural**, **context dependent**, and flexible, while maximising user satisfaction.

Sub-goals

1. Investigate human behaviour given understanding uncertainties.

ightarrow Collect data on possible strategies in WOZ experiment.  ${m {\it O}}$ 

2. Learn a strategy that reflects human behaviour depending on the context.

→ "Bootstrap" an initial policy using SL.

Predicting Multimodal Clarification

Summary & Future work

#### **Thesis Goals**

Overall goal:

We want to learn a clarification strategy which is more **natural**, **context dependent**, and **flexible**, while **maximising user satisfaction**.

Sub-goals

1. Investigate human behaviour given understanding uncertainties.

ightarrow Collect data on possible strategies in WOZ experiment.  ${m {\it O}}$ 

2. Learn a strategy that reflects human behaviour depending on the context.

→ "Bootstrap" an initial policy using SL.

Summary & Future work

# Questions to answer for generating multimodal clarification requests (CRs)

First, the DM needs to decide that "there is evidence of miscommunication" (Gabsdil, 2004). Then, we need to do generation:

#### 1. Content Selection and Organisation

- What level of (mis-) communication to address?
- What severity to indicate?
- 2. Multimodal Output Planning:
  - Uni- or multimodal generation?
- 3. Realisation



Predicting Multimodal Clarification •oo •oo •oo •oo •oo Summary & Future work

#### Outline

Framework Bootstrapping Reinforcement Learning from WOZ Data

## Predicting Multimodal Clarification

#### The Data

Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

(日)
 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

 (1)

## Data Collection: Introducing uncertainties



also see (Skantze, ITRW 03), (Stuttle, ICSLP 04)

Predicting Multimodal Clarification

Summary & Future work

#### The Data

- 24 subjects
- 6 wizards
- 70 dialogues, 1772 turns (774 wizard turns), 17076 words
- 152 Clarification Requests (19.6%)
- 39.5 % multimodal Clarification Requests
- → Can we learn when to generate a multimodal CR in context? (graphic-yes vs. graphic-no)



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Outline

Framework Bootstrapping Reinforcement Learning from WOZ Data

#### Predicting Multimodal Clarification

#### The Data Context/Information-State Features Feature Engineering

Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Local features

- DBmatches: data base matches (numeric)
- deletion: deletion rate (numeric)
- source: problem source (5-valued)
- userSpeechAct: user speech act (3-valued)
- templateGenerated: template generated (binary)
- delay: delay of user reply (numeric)



▲□▶▲□▶▲□▶▲□▶ □□ のQ@

Summary & Future work

## **Dialogue History Features**

- CRhist: number of CRs (numeric)
- screenHist: number screen outputs (numeric)
- delHist: average corruption rate (numeric)
- dialogueDuration: dialogue duration (numeric)
- refHist: number of verbal user references to screen output (numeric)
- clickHist: number of click events (numeric)



Predicting Multimodal Clarification

Summary & Future work

#### User model features

- clickUser: average number of clicks (numeric)
- refUser: average number of verbal references (numeric)
- delUser: average corruption rate for that user (numeric)
- screenUser: average number of screens shown to that user (numeric)
- CRuser: average number of CRs asked to user (numeric)
- driving: user driving (binary)



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Discussion

So far:

- Binary classification task: graphic-yes vs. graphic-no
- 152 training instances
- 19 features, some numeric

How to avoid data sparseness?



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Outline

Framework Bootstrapping Reinforcement Learning from WOZ Data

#### Predicting Multimodal Clarification

The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

## **Discretisation Methods**

"Global discretisation methods divide all continuous features into a smaller number of distinct ranges."

- Unsupervised proportional k-interval discretisation (PKI).
- Supervised/Entropy-based discretisation method based on the Minimal Description Length (MDL) principle.



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

## Feature Selection Methods

"Feature selection refers to the problem of selecting an optimum subset of features that are most predictive of a given outcome."

Searching the feature space:

- forward selection
- backward elimination

Selecting the features:

- Filters:
  - Other ML techniques: J4.8
  - Correlation-based subset evaluation: CFS
  - Correlation-based ranking with cut-off
- Wrappers: Selective Bayes
- Self constructed: Subset overlap



Predicting Multimodal Clarification

Summary & Future work

# Feature selection on PKI-discretised data (left) and on MDL-discretised data (right)



Predicting Multimodal Clarification

Summary & Future work

#### Outline

Framework Bootstrapping Reinforcement Learning from WOZ Data

#### Predicting Multimodal Clarification

The Data Context/Information-State Features Feature Engineering Learning Experiments

Summary & Future work



▲□▶▲□▶▲□▶▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

## **Machine Learners**

Baseline:

- Majority baseline (graphic-no): 45.6 % weighted f-score
- 1-rule baseline: 59.8 % weighted f-score

Machine Learners:

- Rule Induction: RIPPER
- Decision Trees: J4.8
- Naïve Bayes
- Bayesian Network
- Maximum Entropy



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■□ のQ@

Predicting Multimodal Clarification

Summary & Future work

#### Results

Feature transforma-	1-rule	Rule In-	Decision	maxEnt	NB	Bnet	Average
tion/ w. f-score (%)	baseline	duction	Tree				
raw data	59.8	76.1	79.0	76.2	78.5	78.5	74.68
PKI + all features	64.4	72.9	81.6	73.2	81.6	76.4	75.02
PKI+ CFS subset	64.4	75.6	76.3	81.6	81.9***	82.7***	77.08
PKI+ decision tree	64.4	73.8	74.8	81.0	78.9	81.4	75.72
PKI+ selective Bayes	64.4	69.2	74.1	77.9	83.4***	80.0	74.86
PKI+ subset overlap	64.4	76.3	78.5	81.5	83.6***	84.3***	78.10
MDL + all features	69.3	76.9	76.9	79.7	80.4	79.8	77.17
MDL + CFS subset	69.9	76.3	77.2	80.6	81.1	79.8	77.58
MDL + decision tree	75.5	81.5	83.4***	83.4***	83.1***	84.0***	81.82
MDL + select. Bayes	75.5	82.8***	83.4 ***	83.7***	84.1***	84.1***	82.27
MDL + overlap	75.5	82.8***	83.6***	83.6***	84.1***	84.1***	82.28
average	67.95	76.75	78.22	80.78	81.77	81.85	

Predicting Multimodal Clarification

Summary & Future work

## Conclusions

Only the "right" combination of ML model, discretisation method, and feature selection algorithm shows a significant improvement over the 1-rule baseline.

- best performing combinations: Bayesian models with wrapper methods (w. f-score of 84.1%, 58% reduction in error rate)
- MDL discretisation better than PKI.
- 'vertical' differences bigger than 'horizontal'
- best performing feature selection method: subset overlap unversional and the selection method.
- best performing feature subset: templateGenerated, screenHist, screenUser

## Discussion: Best performing feature subset

#### Predictive features:

- + templateGenerated
- + screenHist
- + screenUser
- → Other studies (using RL for feature selection) found repeated concept to be important

Less predictive features:

- refUser
- deletion
- DBmatches
- source
- $\rightarrow$  These (local) features might contribute for a larger data set!



Predicting Multimodal Clarification

Summary & Future work

## Summary

- Framework: "Bootstrap" a RL-based system
- Data collection in a WoZ study.
- Initial strategy learning for when to generate multimodal CRs: 84.1% w. f-score (24.4% improvement over 1-rule baseline)
- Feature engineering as essential step using a large feature space with little data to achieve significant performance gains
- Wizards' behaviour is learnable but is considered to be sub-optimal.

Predicting Multimodal Clarification

Summary & Future work

#### Future work

(Near) future work: Richer annotations

- Add reward level annotations for RL.
- Estimate transition probabilities for MDP for other action decisions (e.g. severity, grounding level).

(Distant) future work:

- Evaluate learnt policy against a hand written strategy.
- Test the portability to other domains.



## Papers associated with this talk:

- Verena Rieser and Oliver Lemon. Learning Multimodal Clarification Strategies: optimizing ISU-based dialogue management from a limited WoZ data-set. Submitted.
- Verena Rieser, Ivana Kruijff-Korbayová, Oliver Lemon.
   Towards Learning Multimodal Clarification Strategies.
   In: 7th ICMI, Doctoral Spotlight, 2005.
- Verena Rieser, Ivana Kruijff-Korbayová, Oliver Lemon: A Framework for Learning Multimodal Clarification Strategies. Proceedings of 6th SIGdial, 2005.



## Weighted f-score

"F-score which says something about recall and precision w.r.t. class frequencies in the data."

$$wf = \sum_{1=1}^{|C|} w_i f(C_i)$$

- Weight the f-score of each class by the class frequency in the data;
- Create the sum .



(日)
 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)
 (日)

 (日)
 (日)

 (日)
 (日)

 (日)

 (日)

 (日)

 (日)
 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

 (日)

#### back

## **Rich Data Annotation**

• <u>Features</u>: Annotation standards for multimodal dialogue context: Joint TALK/AMI workshop, Dec 12th 2005

http://homepages.inf.ed.ac.uk/olemon/standards-workshop-cfp2.html

• <u>Method</u>: NXT format and the NITE XML toolkit (Carletta, 2005)



#### NXT Format



#### NITE toolkit reference coder



#### NITE toolkit gesture coder



#### NITE toolkit dialogue act coder

a Transcription	r 2	Edit Adjacency Pairs	5° 20
Jitendra: okay Statement: <i have="" somethin<="" th=""><th>ng to say about the programming</th><th>Adjacency pair</th><th></th></i>	ng to say about the programming	Adjacency pair	
ractices that are prevalent at IDIAP now. >	ske if even both spread on that	SKIP/ERAC) rome	
litendra, theil things I see now is that everyb	ody has its own code written! is own	source ideas about it.	Influencing-lister
yle.	iouy nus no other code wintern is own		minuencing-incer
Jitendra: but what I feel is that they liking a d	concrete structure.	Type Generic Adjacency Pairs	
Jitendra: so maybe we are programming the	same thing again and again in	En Downou would	
Interent styles different people.	follow come standard while	Target track it.	Uncertain respon
regramming so	TOHOW SOME STATUATO WINE		oncertainrespon
Jitendra: if one person implements something	ng other people don't have to redo it.	New. Delete	
Jitendra: so it's like reusability?? and I know	I'm not talking about something		
ovel because many people		C A edia ana marte Dalara	
Jitendra: thing is being in research organist	Aujacency Pairs		
litendra; at the same time if we can write con	des SKIPIERA C)u usable for future	=)	-
sers maybe it's a good idea.		m4-1.adjacency-pairs.2:	
Jitendra: so Influencing-listeners-action: -	<i a<="" all="" down="" like="" note="" of="" td="" to="" would="" you=""><td> Influencing-listeners-action</td><td></td></i>	Influencing-listeners-action	
KIP(FRAG)s some ideas about it. >		Generic Adjacency Pairs	100
lain: so what did everyone come up with	Information Request: <td>- Oncertain response</td> <td>-</td>	- Oncertain response	-
wek: Negative response: <i agree="" excellently="" td="" wit<=""><td>th that but I I think it's not feasible &gt;</td><td>Edit Dialogue Acts</td><td>S</td></i>	th that but I I think it's not feasible >	Edit Dialogue Acts	S
ecause you know people are so much idealistic! a	and what kind of programming	Con concertatogate Arets	
rogramming platform you want to create I mean	what kind of things are we going to	Dialogue act	
veryone.	where is such in the horizon consult have	Agent: p2	
ack of what has already been written weather you	created a library of the different		
rograms that already exist or howyou would train	ck it. >	DA type: <none></none>	Type
Jitendra: yeah I know but Dialogue-act: <	e certainly know about the people	De text: people working your in ou	A Danas
orking your in our own area. >		own area	· Range
Statements	fication I know who all are	Addressee: p0 p1 p2 0	p3 All
lite Ouestiens			
Questions		NewDA	Deletel
ame. Backchannel and acknowledgment	Acknowledgment		
lite Responses	<ul> <li>Assessment/appreciation</li> </ul>	NITE Video nizver	at al.
Action motivators	• and	I must min	
Jit Checks	maybe you know constructors	Caper-mix	
estruco Politeness mechanisms			
Unlabalad			ynchronise
LITING TO IN L			

#### The End

Thank you for your attention!





<□ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >