

# Learning spatial referential words with mobile robots

Simon Dobnik  
Oxford University

CLUK 06, Milton Keynes  
March 7, 2006

## Spatial expressions

The semantics of spatial expressions

Our goals

Approaches to the semantics of spatial expressions

## Description of our system

Data collection

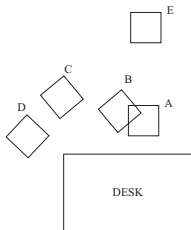
Creating instances for machine learning

Using the acquired knowledge

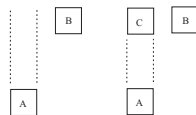
## Conclusion and future work

## Spatial expressions are referential

- ▶ *How near is near? How fast is fast?*
- ▶ We need to evaluate the size of the scene, the perspective at which it is viewed, typical behaviour and properties of objects, and the configuration of other objects (Herskovits, 1986).



(a)



(b)

## Physical world vs. natural language

- ▶ Physical world can be evaluated using *continuous* measures: co-ordinate system with the scale of real numbers.
- ▶ Natural language descriptions use *discrete* reference to refer to events and objects: *near, back, left, slowly, moderately* and *fast*.
- ▶ Non-linguistic reference is made with high degree of accuracy while spatial expressions are ambiguous and vague.

# Aims

The aims of the research are:

- ▶ to learn the meanings of spatial expressions automatically,
- ▶ to be able to demonstrate that the system is able to use them in a way that is natural to a human observer,
- ▶ to integrate the natural language system with the one that is used to drive a mobile robot.

## Why mobile robotics?

- ▶ We get a wealth of information through the robot's sensors (but this information is very low-level). What would be a better way to learn referential expressions?

## Why mobile robotics?

- ▶ We get a wealth of information through the robot's sensors (but this information is very low-level). What would be a better way to learn referential expressions?
- ▶ A robot that can be interacted with in natural language is of great practical utility: interaction with robots in hazardous environments, assistive aids for visually impaired, generating descriptions for virtual environments (computer games), etc.

## Why mobile robotics?

- ▶ We get a wealth of information through the robot's sensors (but this information is very low-level). What would be a better way to learn referential expressions?
- ▶ A robot that can be interacted with in natural language is of great practical utility: interaction with robots in hazardous environments, assistive aids for visually impaired, generating descriptions for virtual environments (computer games), etc.
- ▶ Explore the interaction with the area of mobile robotics that deals with localisation and mapping (SLAM) (Newman, 2001).



## Symbolic vs. non-symbolic

- ▶ Symbolic approaches (Herskovits, 1986; Di Tomaso and Lombardo, 1998) attempt to design rules that encode domain specific knowledge manually.

## Symbolic vs. non-symbolic

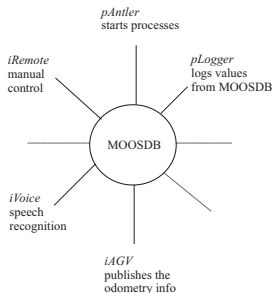
- ▶ Symbolic approaches (Herskovits, 1986; Di Tomaso and Lombardo, 1998) attempt to design rules that encode domain specific knowledge manually.
- ▶ Non-symbolic techniques (Gapp, 1994; Regier and Carlson, 2001) identify abstract parameters that model some properties of physical environment and train their values using machine learning.

## Symbolic vs. non-symbolic

- ▶ Symbolic approaches (Herskovits, 1986; Di Tomaso and Lombardo, 1998) attempt to design rules that encode domain specific knowledge manually.
- ▶ Non-symbolic techniques (Gapp, 1994; Regier and Carlson, 2001) identify abstract parameters that model some properties of physical environment and train their values using machine learning.
- ▶ We follow the second line of research: but we train our classifiers on simple primitives that are available to us through the sensory data of a mobile robot.
- ▶ Resembles the task of *grounding* of word meanings (Roy, 2002).

## MOOS: Mission Oriented Operating Suite

- ▶ A set of libraries and executables that run a mobile robot (Newman, 2001).
- ▶ A modular system with a star-like topology.



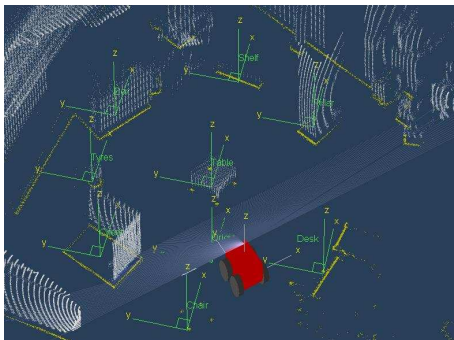
## Data was collected from two contexts

*Context 1:* the robot is moving in an enclosed space and performs a range of motions: forward and backward motion with various velocities, turning left and right under acute and obtuse angles.

- ▶ Describers are asked to describe its motion: *You're going forward slowly. Now you're turning right.*
- ▶ All descriptions are made from the robot's point of view.
- ▶ The robot is controlled by an operator who attempts to go through a range of motions for each describer but in no particular order.
- ▶ Considerable errors: a delay after a description was made and before it reached the MOOS database!

## Data was collected from two contexts

*Context II:* static scenes with real-size objects such as desks, chairs and walls and the describers comment on the locations of objects (including the robot): *The chair is to the left of you. The table is further away than the chair.*



## Structure of datasets for machine learning

- ▶ Weka toolkit (Witten and Frank, 2000) is used.
- ▶ A supervised method: information must be preprocessed and abstracted in a certain way (may effect the learning procedure).
- ▶ Datasets must consist of independent instances which are vectors of attribute values, the properties that we want to include in learning.
- ▶ To learn the value of an attribute (also known as the target concept) means to find a relationship between other attribute values.

## From MOOS logs to Weka input files

MOOS log files look like this (simplified):

```
24.452 ODOMETRY ..., dh=0.635, speed=0.542, ...
24.578 ODOMETRY ..., dh=0.000, speed=0,121, ...
24.623 COMMENTARY_RELATIONS ..., Desk=[3x1]{-1.916,5.136,0}...
24.649 ODOMETRY ..., dh=0.001, speed=0.234, ...
25.034 VOICE_INPUT You're turning left.
```

For Weka we need something like this:

```
0.001, 0.234, turning, left, none, none
```

Some numeric values must be calculated. All of them must be normalised. The category membership of words must be determined (simple unification grammar).



## And finally. . .

- ▶ *Context I*:  $\langle \text{delta\_heading, speed, verb, direction, heading, manner} \rangle$
- ▶ *Context II*:  $\langle \text{lo\_x, lo\_y, refo\_x, refo\_y, relation} \rangle$

## Decision trees

- ▶ Select an attribute to create a node for in the tree and then create branches for each of its possible values.
- ▶ Repeat the process (using a subset of instances that fall under that branch) until all instances at a node have the same classification.
- ▶ Prefers small trees: uses *information gain* as the measure to choose which attribute to split on first.

# NaiveBayes

- ▶ A rule generator based on Bayes' rule of conditional probability.
- ▶ What is the probability of a description (the target class) given some evidence (the state of the robot and the environment)?
- ▶ 
$$\Pr(\text{desc} | E_{1...n}) = \frac{\Pr(E_1 | \text{desc}) \dots \Pr(E_n | \text{desc}) \Pr(\text{desc})}{\Pr(E)}$$
- ▶ Find the probabilities on the RH of the equation.
- ▶ All attributes are equally important and independent of one another.

## Context I: datasets

- ▶ *Subset I*: 192 instances from 'the best dataset' with no alignment.
- ▶ *Subset II*: 338 instances created with alignment.
- ▶ When learning a description, only numeric attributes were included. For example:  $\langle \text{delta\_heading}, \text{speed}, \text{verb} \rangle$ .
- ▶ 10-fold cross-validation was used to test the accuracies of the classifiers.

## Context I: results for nominal attributes

### Subset I

<i>Classifier</i>	<i>Direction</i>	<i>Heading</i>	<i>Manner</i>	<i>Verb</i>
Decision trees	74.0%	74.0%	79.2%	75.5%
NaiveBayes	67.7%	75.5%	78.1%	64.1%

### Subset II

<i>Classifier</i>	<i>Direction</i>	<i>Heading</i>	<i>Manner</i>	<i>Verb</i>
Decision trees	73.4%	73.1%	65.4%	65.4%
NaiveBayes	70.1%	67.8%	57.4%	62.1%

## Context I: which words were learnt?

### Nominal attribute values for Subset I

<i>Attribute</i>	<i>Values</i>
Direction	backward, forward, none, spot, straight
Heading	anticlockwise, clockwise, left, none, right
Manner	fast, moderately, none, slowly
Verb	creeping, going, moving, turning, stopped

Since each attribute has 4 or 5 values, the probability of randomly guessing a word for each class is 25% or 20%.

## Context I: continuous numeric attributes

### The accuracy of decision trees for Subset I

<i>Bins</i>	<i>Delta heading</i>	<i>Speed</i>
3	81.8%	80.2%
5	72.4%	69.8%
10	49.5%	64.0%
20	35.9%	52.6%
30	30.7%	52.1%
40	26.0%	49.5%

## Context II: prepositional relations

- ▶ 251 instances in total.
- ▶  $\langle lo\_x, lo\_y, refo\_x, refo\_y, relation \rangle$
- ▶ 10-fold cross-validation.
- ▶ Correctly classified instances: 74.9% (Decision trees) and 77.3% (NaiveBayes).



# pDescriber

- ▶ *pDescriber* is a commentator.
- ▶ If the robot is moving, it describes its actions: *I'm going forward fast.*
- ▶ If the robot is stationary, it provides comments about position of objects: *The table is to the right of the chest.*
- ▶ Uses nominal classifiers for both contexts and the unification grammar to generate sentences.
- ▶ The system includes a speech synthesiser.

## pDialogue

- ▶ *Chats with users*: a pattern matching dialogue interface that matches user's input with a predefined pattern and returns the associated reply.
- ▶ *Performs motion commands*: *Go forward slowly. Go forward right fast.* It uses the classifiers for *Delta heading* and *Speed*, turns them to *Desired rudder* and *Desired thrust* (commands how to achieve that state).
- ▶ *Answers questions about position of objects*: *Where is the chair?* It uses the nominal classifier for relations.
- ▶ In both (2) and (3) our simple unification grammar is used to generate sentences or to parse the input.

## Conclusion and future work

- ▶ A complete cycle how descriptions are learnt from the properties of the environment internalised by the robot and subsequently used to refer to the environment.
- ▶ Minimise human input in supervised learning.
- ▶ Perform learning on higher (abstracted) levels of robot's environment (modelling of perspective, etc.).
- ▶ Evaluate and minimise errors in input data.
- ▶ How can linguistic and localisation (SLAM) system be integrated to benefit from each other?