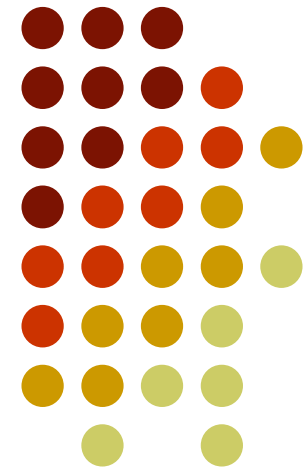


Communication Skills

Jesse Davis

Dept. of Computer Science
Katholieke Universiteit Leuven





Motivation

In order to graduate you must

- Write a thesis
- Orally defend your thesis

How well you convey your ideas has a significant impact your grade



PART I: WRITING



Outline

- Fundamentals of writing
- Structuring a thesis
- General advice and common errors
- Plagiarism
- Exercises



Writing Commandments

START WRITING EARLY

Writing should be

- Clear
- Precise
- Non-redundant
- Consistent

Start Writing Early



- Writing takes a long time
- Paragraphs or chapters often must be rewritten many times
- Hard to write, even for native speakers

Each Sentence Should Be Clear



- Does a sentence make sense?
- Does it clearly convey your point?
- Does it contain all the necessary information?
- Examples:
 - Handling an unobserved strategy can be solved using the EM scheme.
 - The learning algorithm does not allow for any hidden information except the selection.



Write Concisely

- Do you need to every word in the sentence?
- Is there a shorter way to express your point?
- Example:
 - **Wordy:** “Because one customer has bought a pair of books together, every other customer that is interested in one of these books is not necessarily also interested in the other one.”
 - **Concise:** “Just because one customer bought a pair of books does not imply that every other customer that buys one of these books will buy the other book.”

Non-Redundant



- The main messages should be repeated, but that is it
- Look at paragraph/sections and see if the same information appears more than once
- Ties back to being concise

BE CONSISTENT



- Use the same terminology through the paper
 - E.g.: Switching between SRL, PLL would be confusing
- Use the structure consistently
 - E.g., italicize a definition
 - E.g., different fonts for formulas

Tips for Successful Writing



- Start writing early
- Break up the text with figures/lists/different environments
- Rewrite liberally: I go through many many iterations before I'm happy
- Start writing early!



Outline

- Fundamentals of writing
- Structuring a thesis
- General advice and common errors
- Plagiarism
- Exercises



Paper Structure

- Introduction
- Background and related work
- Chapters on contributions
- Conclusions and future work

What to Convey in a Thesis



- What is the problem you are solving and why is it important and challenging
 - Provide motivation and context for the work
- What is your solution to the problem
 - How did you solve the problem
 - How is your solution different
 - What is the value added of your approach

What to Convey in a Thesis: Introduction



Provide motivation and context

- What is the problem you are solving
- Why is it important and challenging
- Where is your work situated in the field
- How your solution is different
- Brief overview of what you did

What to Convey in a Thesis: Background



Set the scene

- Define necessary terminology
- Provide knowledge the reader needs to understand the thesis
- Compare and contrast your thesis to existing work

What to Convey in a Thesis: Your Contributions



- Tell the reader what you did
 - Describe your solution
 - Provide enough details to redo your work
 - Justify decisions you make
- Experiments
 - Start with a question you want to answer
 - Describe how you will evaluate it
 - Think about appropriate baselines algorithms

What to Convey in a Thesis: Conclusions



Wrap up the thesis

- Highlight the main contributions
 - Algorithmically
 - Empirically
- Focus on the take away message: What should the reader have learned?
- Briefly mention what the next steps may be



Outline

- Fundamentals of writing
- Structuring a thesis
- General advice and common errors
[Look at Marie Des Jardin's Web page]
- Plagiarism
- Exercises



General Rules

- Avoid contractions
 - Use “does not” instead of “doesn’t”
 - Use “it is” instead of “it’s”
- Spell out numbers less than or equal to ten
- Avoid transitioning from a section directly to a subsection without intervening text
- Avoid a section that has only one subsection
- Avoid colloquial English or slang



Be Parallel: Verb vs. Noun

- Bad: The system involves
 - Knowledge extraction
 - Inferring novel facts
 - Model construction
- Good: The system involves
 - Extracting knowledge
 - Making novel inferences
 - Learning predictive models



Be Parallel: Verb Form

- Bad: The system works by
 - Extracting knowledge
 - Infer novel facts
 - Learn predictive model
- Good: The system involves
 - Extracting knowledge
 - Making novel inferences
 - Learning predictive models



Be Parallel: Listing Points

- Bad:
 - First of all,
 - Second,
 - Thirdly,
- Good:
 - First,
 - Second,
 - Third



Avoid the Passive Voice

- Use the verb “to be” for definitions
 - Learning is the ability to improve with experience
 - X is
- Avoid things like:
 - The following example is used to illustrate
 - The algorithm is supposed
 - The full state is generated from a set of variables
 - Only temporal sequences are considered

Do No Repeat Words in Sentences/Paragraphs



- Think about word choice, use dictionary and thesaurus if needed
- Example: “Exact inference in hybrid models is **prohibitively** slow so a learning method using exact estimations of the marginal distributions is **prohibitive**.”
- Note: This cannot always be avoided



Common Errors

- “i.e.,” means “that is” and must be written “i.e.”
- “e.g.,” means “for example” and must be written “e.g.”
- “Et al.”
- Footnotes go after periods “.\footnote{”
- Punctuation goes inside quotes
 - E.g., “This is correct.”



Common Errors

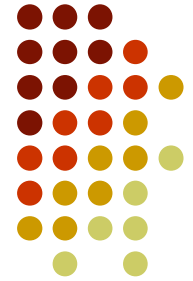
- “its” is possessive
- “it’s” means it is
- Citations cannot appear as nouns!
 - **Bad:** “(Davis et al., 2005) shows”
 - **Bad:** “[17] shows”
 - **Good:** “Davis et al. (2005) shows”
- Between is used for only two entities
- Among is used for more than two entities



Outline

- Fundamentals of writing
- Structuring a paper
- General advice and common errors
- Plagiarism
- Exercises

Copying Text Word for Word from Another Source



- You must cite the other work!!!!
- Failing to do this constitutes fraud
- Note that employing direct quotations should rarely (if ever) appear in a MAI thesis
 - If one occurs, it should be set in quotation marks
 - If the quotation is multiple sentences, it should be indented from the rest of the text

Paraphrasing Text from Another Source



- You must cite the other work!!!!
- Failing to do this constitutes fraud
- Examples of this may include, rephrasing the descriptions of ML algorithms from ML course text book
- Again, I would avoid paraphrasing in a thesis

Copying Tables and Figures from Another Source



- You must cite the other work!!!!
- Include a citation both
 - In the text of the thesis when the table, figure, image, etc. is described
 - In the caption of the table, figure, image, etc.



Outline

- Fundamentals of writing
- Structuring a paper
- General advice and common errors
- Plagiarism
- Exercises



Teach vs. Learn

- I _____ myself how to play the piano
- I _____ how to play the piano
- I _____ about machine learning in class
- Luc _____ about probabilistic models in class
- Luc _____ probabilistic models in class



Remove Passive Voice

- The key points are illustrated in Figure 1.
- The company is based in Portland and designs microprocessors.
- The decisions that are selected should maximize the expected utility.
- Variables are used as placeholders for specific entities.
- The family of probabilistic programming languages considered in this thesis are based on logic programming.



PART II: PRESENTATIONS



Outline

- General advice for presentations
- Improving readability
- Using pictures
- Presenting results
- Tips for making slides

Goals of Presenting



- Goal is to give big picture, not all the details
- Think about: What is the take away message
- Focus on the thesis' most important aspects



General Speaking Tips

- Look at the audience
 - Do not speak to the screen
 - Try to keep eye contact
- Speak slowly
- Be energetic
- Practice...a lot



General Slide Advice

- Think about the structure of the talk
- Make sure to introduce all relevant background knowledge/terms
- Think about how to best convey ideas
- Making good slides takes a long time

Structuring a Presentation



Like writing, want to convey

- What is the problem
- Why is it important
- How did you solve it
- What results did you get

Defining Your Task



- Be sure to clearly define the problem you are trying to solve
- A good strategy is “Given” and “Do”
- Another good strategy is to show a picture

Simplified Clinical Database



Patients

PID	Gender	Birthday
P1	M	3/22/63

Prescriptions

PID	Prescribed	Medication	Dose	Duration
P1	5/17/98	prilosec	10mg	3 months

Diseases

PID	Date	Symptoms	Diagnosis
P1	1/1/01	palpitations	hypoglycemic
P1	2/1/03	fever, aches	influenza

Lab Tests

PID	Date	Lab Test	Result
P1	1/1/01	blood glucose	42
P1	1/9/02	blood glucose	45

Goal: Predict at prescription time if a patient will have an adverse reaction to a medicine



Outline

- General advice for presentations
- Improving readability
- Using pictures
- Presenting results
- Tips for making slides



**A PRESENTATION IS FOR THE
AUDIENCE**

MAKE IT EASY ON THEM!

Colors and Spacing Can Greatly Improve Readability



- Spaces between points is helpful
- It makes things easier to read for a listener
- Color highlights important points



Contributions of this Talk

- Real-world problems violate standard assumptions
 - Complex and uncertain data
 - Implicitly defined features and relations
 - Train and test data come from different distributions
- My work: Discover probabilistic, structural regularities
 - Invent predicates/features: More accurate learned models
 - Abstract predicates: Reuse knowledge across domains
- Large improvements on significant applications
- Many remaining technical challenges!



Contributions of this Talk

- Real-world problems violate standard assumptions
 - Complex and uncertain data
 - Implicitly defined features and relations
 - Train and test data come from different distributions
- My work: Discover probabilistic, structural regularities
 - Invent predicates/features: More accurate learned models
 - Abstract predicates: Reuse knowledge across domains
- Large improvements on significant applications
- Many remaining technical challenges!



Other Readability Issues

- In terms of colors, difficult to see
 - Dark on dark
 - Light on light
 - Think about people who are color-blind
- Try to use at least 24 point font
- Think about people in the back of the room
- Make sure top, bottom, sides not cut off



Outline

- General advice for presentations
- Improving readability
- Using pictures
- Presenting results
- Tips for making slides



**A PRESENTATION IS NOT A
COLLECTION OF BULLET POINTS**

USE PICTURES!!!!

Classifying Unseen Data



- Proposition methods treat each example independently
- Predict label for each example using only the attribute-values for that example

Classifying Unseen Data



Abnormality	Patient	Date	Calcification Fine/Linear	...	Mass Size	Loc	Cancer
1	P1	5/02	Absent		3mm	RU4	?
2	P1	5/04	Present		5mm	RU4	?
3	P1	5/04	Absent		4mm	LL3	?
4	P2	6/00	Absent		2mm	RL2	?
...

Classifying Unseen Data



Abnormality	Patient	Date	Calcification Fine/Linear	...	Mass Size	Loc	Cancer
1	P1	5/02	Absent		3mm	RU4	No
2	P1	5/04	Present		5mm	RU4	?
3	P1	5/04	Absent		4mm	LL3	?
4	P2	6/00	Absent		2mm	RL2	?
...

Bayesian Networks

[Pearl, 1988]

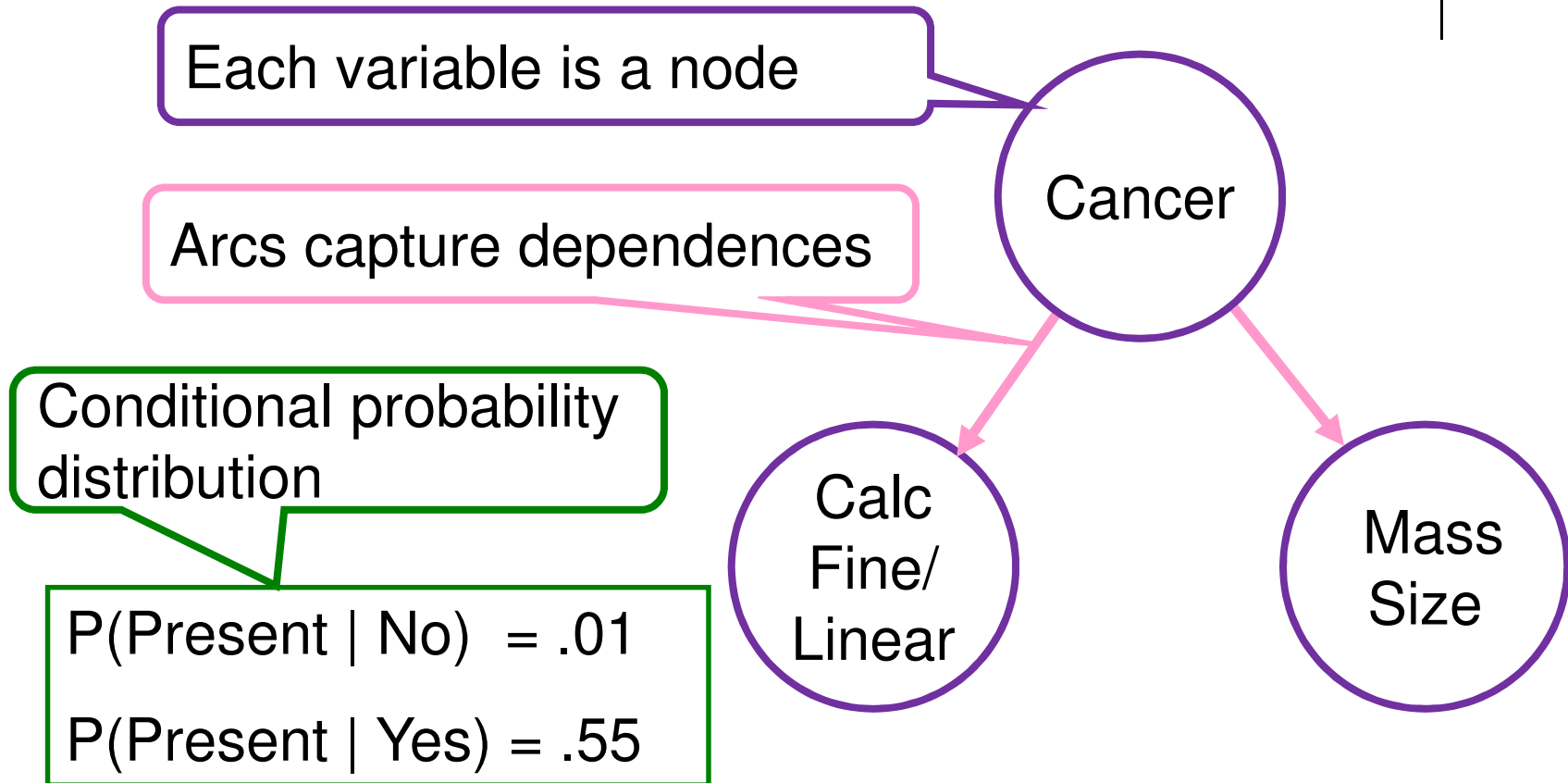


- Directed, probabilistic graphical model
 - Nodes = variables
 - Arcs = probabilistic dependency between nodes
- Each node x has a conditional probability distribution: $\text{Prob}(x \mid \text{Parents}(x))$
- Encodes the following distribution

$$P(x) = \prod_i P(x_i \mid \text{Parents}(x_i))$$

Bayesian Networks

[Pearl, 1988]



Joint Distribution:
$$P(x) = \prod_i P(x_i \mid Parents(x_i))$$



Outline

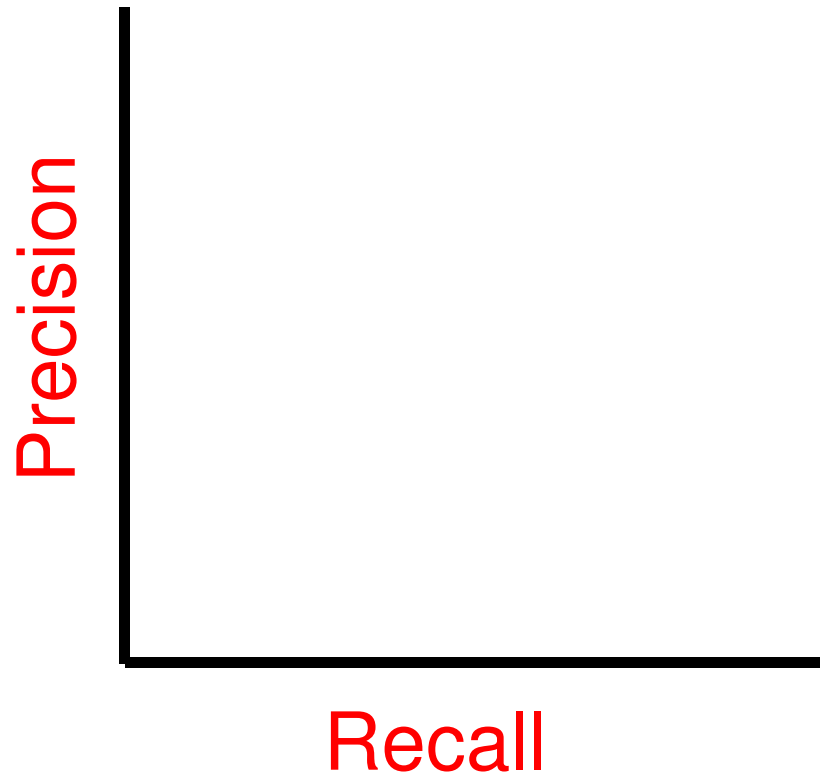
- General advice for presentations
- Improving readability
- Using pictures
- Presenting results
- Tips for making slides



Presenting Results

- What is the point of the experiments?
- What datasets were used?
- What are the relevant metrics?
- What are the relevant baselines?

Presenting Results



- Label axes
- Describe axes and what plot shows
- Describe what good/best results looks like



Presenting results

- Tables are too hard to read

	Task 1	Task 2	Task 3	Task 4
Algo 1	0.6	0.45	0.58	0.71
Algo 2	0.7	0.55	0.82	0.71
Algo 3	0.75	0.5	0.83	0.72

- Take the time and make a chart



Outline

- General advice for presentations
- Improving readability
- Using pictures
- Presenting results
- Tips for making slides

Slide Tips:

Avoid Single Bullet Points

- This is pointless



Slide Tips:

Avoid Single Bullet Points



- Yada yada
 - This looks silly
- Blah blah
- Blah blah

Slide Tips: Do Not Over Nest Bullet Points



- You need at most
 - Two levels of points
 - Points on the third level are bad

Slide Tips: Avoid Bad Line Breaks



Bad

- Proposes new feature as first-order definite clauses
- Introduce feature as binary variable in statistical model

Good

- Proposes new feature as first-order definite clauses
- Introduce feature as binary variable in statistical model



Presentation Wrap Up

Avoid the following pit falls:

- Lacking a take away message
- Not practicing the presentation enough
- Including too much text on the slides
- Allocating insufficient time for making slides



Take Away

- Communication skills are very important and you need to practice them
- Make sure to devote lots of time to
 - Writing
 - Preparing presentations
- Start early and ask for feedback



EXTRA EXERCISES



Make More Concise

- Many systems are based on the observation that existing probabilistic models can not cope with the rich relational representations but can be “upgraded”. This upgrading approach starts from well-known propositional systems, where efficient inference algorithms are available, and extends the representation by relational concepts.



Make More Concise

- Traditional machine learning techniques expects data to come in the form of feature vectors. A feature vector describes the training example by a set of features.
- The second skill is the ability to **decide** what to do, and is concerned with deciding which actions to take.

Make More Precise & Concise



A second motivation for our work is due to the observed differences of domains these models are typically applied to. It can be observed that the approaches not only differ in representations but also in the way inference is performed and therefore domains they are applied to. Some use knowledge-based model construction, which compiles the query and/or evidence into a propositional graphical model like Bayesian or Markov network. These graphical model representations allow one to apply a wide variety of existing inference and learning algorithms, which allows one to use them for many applications.

Make Parallel



This learning algorithm is closely related to the one for CPT-L. First of all it provides a motivation for the propositional logical formula generated by CPT-L, which can be seen as specialization of Clark's completion. Furthermore, it generalizes EM by means of the BDD algorithm of CPT-L. The generalization allows for hidden and deterministic variables. Third, as in CPT-L it splits sequences into transitions; LFI-ProbLog exploits certain Independence to split training examples. In fact, LFI-ProbLog, when applied to translated CPT-L, would automatically rediscover its learning algorithm. Finally, very similar to the partial lifted algorithm of CPT-L, LFI-ProbLog provides a probabilistic version of unit propagation.