Introduction to

Artificial Intelligence

with Python

Knowledge

Goal : building knowledge-based agents

knowledge-based agents

agents that reason by operating on internal representations of knowledge

Reasoning-based knowledge :

If it didn't rain, Harry visited Hagrid today.

Harry visited Hagrid or Dumbledore today, but not both.

Harry visited Dumbledore today.

Harry did not visit Hagrid today.

It rained today.

Logic

sentence

an assertion about the world
in a knowledge representation language

Propositional Logic

Propositions = some statements about world

Proposition Symbols

P Q R

Logical Connectives

$\neg$	$\wedge$	$\vee$
not	and	or

$\rightarrow$	$\leftrightarrow$
implication	biconditional

Not ($\neg$) (essentially opposites)

P	$\neg P$
false	true
true	false

And (∧)

P	Q	$P \wedge Q$
false	false	false
false	true	false
true	false	false
true	true	true

Or (v)

as long as at least 1 operand is True, evaluates to True. Only need either 1 or both to be True.

P	Q	$P \vee Q$
false	false	false
false	true	true
true	false	true
true	true	true

Implication ($\rightarrow$)

if P is t, Q also needs to be t → otherwise P is f

"P implies Q"

no actual claim being made about Q

P	Q	$P \rightarrow Q$
false	false	true
false	true	true
true	false	false
true	true	true

Biconditional ($\leftrightarrow$)

Condition goes in both directions
e.g. I'll be indoors if and only if
it is raining

P	Q	$P \leftrightarrow Q$
false	false	true
false	true	false
true	false	false
true	true	true

Tells us whether things are actually true in world.

model

assignment of a truth value to every propositional symbol (a "possible world")

model

P: It is raining.

Q: It is a Tuesday.

$\{P = \texttt{true}, Q = \texttt{false}\}$ → these variables will have diff values in diff. models

knowledge base

a set of sentences known by a knowledge-based agent i.e. things that AI essentially knows to be true

Entailment

"α entails β"

$$\alpha \vDash \beta$$

In every model in which sentence **α** is true, sentence β is also true.

$\therefore$ if α is true, β must also be True

We essentially want AI to find entailments.

If it didn't rain, Harry visited Hagrid today.

Harry visited Hagrid or Dumbledore today, but not both.

Harry visited Dumbledore today.

Harry did not visit Hagrid today.

It rained today.

inference

the process of deriving new sentences from old ones

P: It is a Tuesday.

Q: It is raining.

R: Harry will go for a run.

It's Tuesday and not rain then H will run.

KB: $(P \land \neg Q) \rightarrow R$ $\quad P \quad \neg Q$

we know this is true

we know this is true

Inference: R

Inference Algorithms

Q asked by inference algorithms?

Does

$$\mathrm{KB} \models \alpha$$

?

Model Checking

= enumerates truth values

think of these as possible worlds : some assignment of truth values

Model Checking

- To determine if $\mathbf{KB} \vDash \alpha$:
  - Enumerate all possible models/worlds,
  - If in every model where $\mathbf{KB}$ is true, α is true, then $\mathbf{KB}$ entails α.
  - Otherwise, $\mathbf{KB}$ does not entail α.

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

KB: $(P \wedge \neg Q) \rightarrow R$ P $\neg Q$

Query: R (α)

8 poss ways to assign t/f to our 3 propositional symbols

we know P is true

P	Q	R	KB
false X	false	false	F
false X	false	true	F
false X	true X	false	F
false X	true X	true	F
true	false	false	F
true	false	true	T
true	true X	false	F
true	true X	true	F

we know Q is false

1 world when KB is true and R is true
∴ we can draw inference

Q: in which worlds is KB true?

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

KB: $(P \wedge \neg Q) \rightarrow R$ P $\neg Q$

this is false because the $(P \wedge \neg Q) \rightarrow R$ implication is also part of KB

Query: R

P	Q	R	KB
false	false	false	false
false	false	true	false
false	true	false	false
false	true	true	false
true	false	false	false
true	false	true	true
true	true	false	false
true	true	true	false

P: It is a Tuesday. Q: It is raining. R: Harry will go for a run.

KB: $(P \land \neg Q) \rightarrow R$ P $\neg Q$

Query: R

Only 1 world here where KB is true. Here, R is actually so we can draw the conclusion.

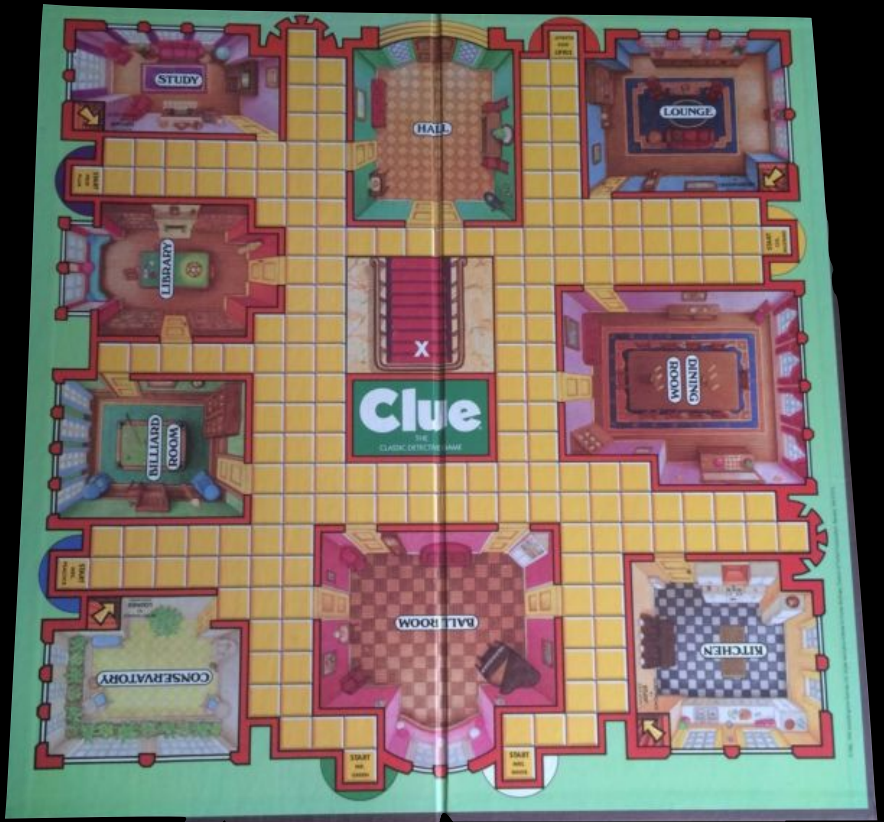
P	Q	R	KB
false	false	false	false
false	false	true	false
false	true	false	false
false	true	true	false
true	false	false	false
true	false	true	true
true	true	false	false
true	true	true	false

Knowledge Engineering

Taking a problem and breaking it into knowledge that's distillable by computer

Look at examples...

Clue



STUDY
HALL
LOUNGE
LIBRARY
DINING ROOM
BILLIARD ROOM
Clue
BALLROOM
KITCHEN
CONSERVATORY

Clue

People	Rooms	Weapons
Col. Mustard	Ballroom	Knife
Prof. Plum	Kitchen	Revolver
Ms. Scarlet	Library	Wrench

Clue

People

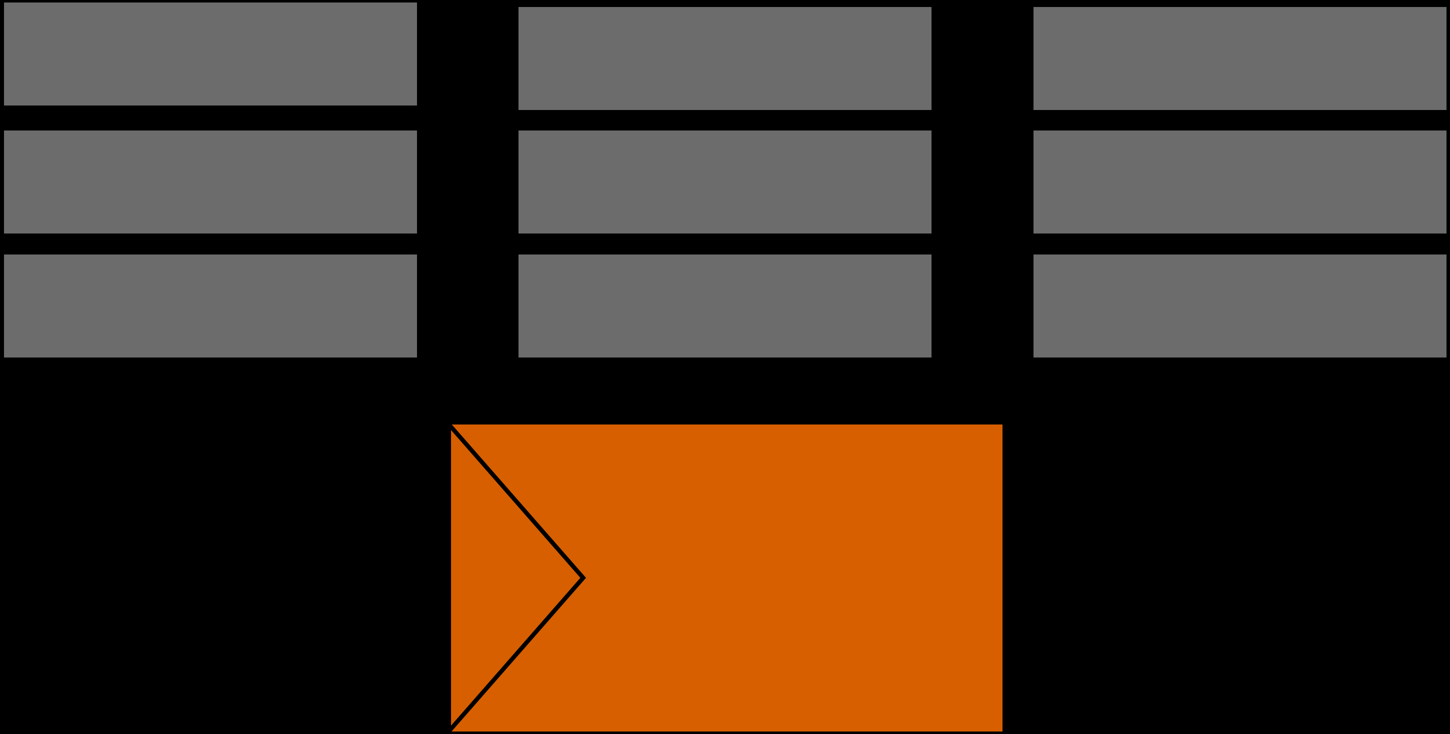
Rooms

Weapons

Clue

People	Rooms	Weapons

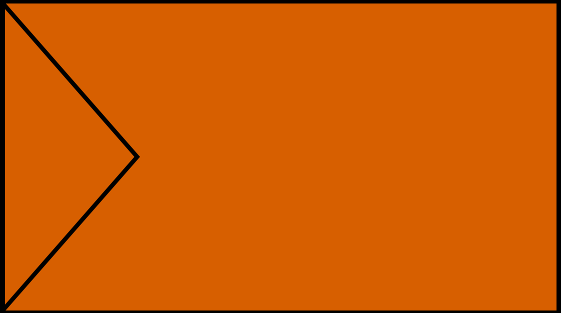
People Rooms Weapons



People

Rooms

Weapons



Look at some cards to figure out what's going on.

Clue

These will be T if
they are in envelope

Propositional Symbols

mustard	*ballroom*	*knife*
plum	*kitchen*	*revolver*
scarlet	*library*	*wrench*

Clue

we know one of these 3 ppl is the murderer

$$(mustard \lor plum \lor scarlet)$$

One of the rooms must be in envelope

$$(ballroom \lor kitchen \lor library)$$

one of those weapons is in envelope

$$(knife \lor revolver \lor wrench)$$

if have plum in hand, know he's not in envelope

$$\neg plum$$

$$\neg mustard \lor \neg library \lor \neg revolver$$

know that at least one of them not in envelope

Logic Puzzles

- Gilderoy, Minerva, Pomona and Horace each belong to a different one of the four houses: Gryffindor, Hufflepuff, Ravenclaw, and Slytherin House.
- Gilderoy belongs to Gryffindor or Ravenclaw.
- Pomona does not belong in Slytherin.
- Minerva belongs to Gryffindor.

Logic Puzzles

1 for each person/house combo

Propositional Symbols - each is either T or F

GilderoyGryffindor
GilderoyHufflepuff
GilderoyRavenclaw
GilderoySlytherin

MinervaGryffindor
MinervaHufflepuff
MinervaRavenclaw
MinervaSlytherin

PomonaGryffindor
PomonaHufflepuff
PomonaRavenclaw
PomonaSlytherin

HoraceGryffindor
HoraceHufflepuff
HoraceRavenclaw
HoraceSlytherin

Logic Puzzles

$$(PomonaSlytherin \rightarrow \neg PomonaHufflepuff)$$

because each person in only 1 house

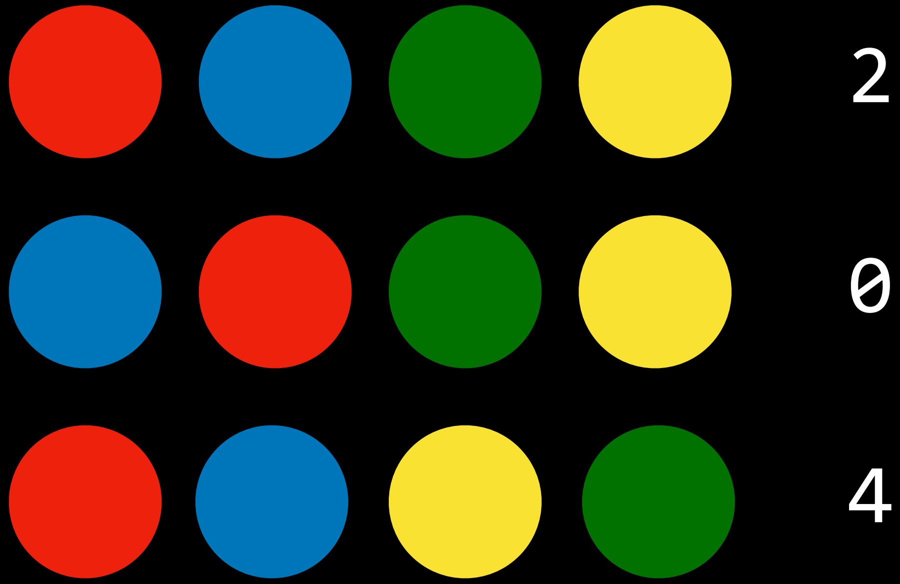
$$(MinervaRavenclaw \rightarrow \neg GilderoyRavenclaw)$$

because each person in diff. house

$$(GilderoyGryffindor \vee GilderoyRavenclaw)$$

Mastermind

N.B. that mastermind.py runs slow → model checking not the most efficient (need to enumerate all possibilities → 2^N if we have N variables)



Need better way to make inferences

Inference Rules

= rules to take existing knowledge and transform into new knowledge

Modus Ponens

If it is raining, then Harry is inside.

kb we know ← It is raining.

---

inferences we can make ← Harry is inside.

Modus Ponens

if we know that α implies β
AND
we know α is true, then we also know β is true

N.B. not dealing with any specific worlds!

$$\alpha \to \beta$$

$$\frac{\alpha}{\beta}$$

And Elimination

Harry is friends with Ron and Hermione.

Harry is friends with Hermione.

And Elimination

if α and β then just α and just β must also be true

Obvious but need to tell computer

$$\frac{\alpha \wedge \beta}{\alpha}$$

Double Negation Elimination

It is not true that Harry did not pass the test.

Harry passed the test.

Double Negation Elimination

$$\frac{\neg(\neg\alpha)}{\alpha}$$

If we have two negatives, they cancel each other out.

Implication Elimination

One of the two must be true

If it is raining, then Harry is inside.

It is not raining or Harry is inside.

Implication Elimination

Translating 'if' into 'or' statements.

$$\alpha \rightarrow \beta$$

If I have this implication

$$\neg\alpha \vee \beta$$

Have to conclude that either not α or β

Biconditional Elimination

It is raining if and only if Harry is inside.

Implication goes both ways

If it is raining, then Harry is inside,
and if Harry is inside, then it is raining.

Biconditional Elimination

$$\frac{\alpha \leftrightarrow \beta}{(\alpha \rightarrow \beta) \land (\beta \rightarrow \alpha)}$$

De Morgan's Law

turning this "and" ~~in~~ into an "or"

It is not true that both
Harry and Ron passed the test.

Harry did not pass the test
or Ron did not pass the test.
(or both)

De Morgan's Law

$$\frac{\neg(\alpha \land \beta)}{\neg\alpha \lor \neg\beta}$$

Moving the negation inwards

De Morgan's Law

Reverse of De Morgan's law

turning the "or" into an "and"

It is not true that (Neither passed the
Harry or Ron passed the test. test)

Harry did not pass the test
and Ron did not pass the test.

De Morgan's Law

$$\frac{\neg(\alpha \vee \beta)}{\neg\alpha \wedge \neg\beta}$$

Distributive Property

Like distributive laws in maths

$$\frac{(\alpha \wedge (\beta \vee \gamma))}{(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)}$$

Distributive Property

$$\frac{(\alpha \vee (\beta \wedge \gamma))}{(\alpha \vee \beta) \wedge (\alpha \vee \gamma)}$$

Q: how can we use inference rules to draw conclusions?

Search Problems

1. Can treat our sets of sentences as states in our search problem.

- initial state
- actions → Inference rules are actions
- transition model → tells us what we get from inference rules
- goal test → check if thing we're trying to prove
- path cost function → num of inference rules we need to apply

Theorem Proving

- initial state: starting knowledge base
- actions: inference rules
- transition model: new knowledge base after inference
- goal test: check statement we're trying to prove
- path cost function: number of steps in proof

Resolution = powerful inference rule

Complimentary literals that conflict w/ each other

(Ron is in the Great Hall) v (Hermione is in the library)

Ron is not in the Great Hall

Hermione is in the library

Can start to generalise quickly eg. multiple Qs

$$P \vee Q$$

$$\neg P$$

$$Q$$

$$P \vee Q_1 \vee Q_2 \vee \ldots \vee Q_n$$

$$\neg P$$

---

resolved into:

$$Q_1 \vee Q_2 \vee \ldots \vee Q_n$$

Can generalise further...

(Ron is in the Great Hall) v (Hermione is in the library)

because these conflict

either of these must be true

(Ron is not in the Great Hall) v (Harry is sleeping)

(Hermione is in the library) v (Harry is sleeping)

$$P \lor Q$$

$$\neg P \lor R$$

$$Q \lor R$$

Q & R don't need to be singu propositional symbols

$$P \lor Q_1 \lor Q_2 \lor \ldots \lor Q_n$$

$$\neg P \lor R_1 \lor R_2 \lor \ldots \lor R_m$$

$$Q_1 \lor Q_2 \lor \ldots \lor Q_n \lor R_1 \lor R_2 \lor \ldots \lor R_m$$

things connected with "or"

clause

a disjunction of literals

either a propositional symbol or a negation of a proposit. symbol

e.g. $P \vee Q \vee R$

allow us to turn sentences into

things connected by "and"

conjunctive normal form

logical sentence that is a conjunction of clauses

e.g. $(A \vee B \vee C) \wedge (D \vee \neg E) \wedge (F \vee G)$

Can turn any sentence into CNF using inference rules

Conversion to CNF

- Eliminate biconditionals
  - turn $(\alpha \leftrightarrow \beta)$ into $(\alpha \rightarrow \beta) \land (\beta \rightarrow \alpha)$
- Eliminate implications
  - turn $(\alpha \rightarrow \beta)$ into $\neg\alpha \lor \beta$
- Move ¬ inwards using De Morgan's Laws
  - e.g. turn $\neg(\alpha \land \beta)$ into $\neg\alpha \lor \neg\beta$
- Use distributive law to distribute ∨ wherever possible

Conversion to CNF

$(P \vee Q) \rightarrow R$

$\neg(P \vee Q) \vee R$ eliminate implication

$(\neg P \wedge \neg Q) \vee R$ De Morgan's Law

$(\neg P \vee R) \wedge (\neg Q \vee R)$ distributive law

conjunctive normal form (conjunction of disjunctions)

We care because can resolve this using resolution!

Inference by Resolution

$$P \vee Q$$

$$\neg P \vee R$$

$$(Q \vee R)$$

$$P \vee Q \vee S$$

$$\neg P \vee R \vee S$$

$$(Q \vee \underline{S} \vee R \vee \underline{S})$$

double s redundant → eliminate
(factoring)

$$\frac{\begin{array}{c} P \vee Q \vee S \\ \neg P \vee R \vee S \end{array}}{(Q \vee R \vee S)}$$

$$\frac{\begin{array}{c} P \\ \neg P \end{array}}{(\,)}$$

→ emty clause = always false

Inference by Resolution

common technique in CS

- To determine if $\mathbf{KB} \vDash \alpha$:
  - Check if $(\mathbf{KB} \land \neg\alpha)$ is a contradiction?
    - If so, then $\mathbf{KB} \vDash \alpha$.
    - Otherwise, no entailment.

Inference by Resolution

- To determine if $\text{KB} \vDash \alpha$:
  - Convert $(\text{KB} \wedge \neg\alpha)$ to Conjunctive Normal Form.
  - Keep checking to see if we can use resolution to produce a new clause.
    - If ever we produce the **empty** clause (equivalent to False), we have a contradiction, and $\text{KB} \vDash \alpha$.
    - Otherwise, if we can't add new clauses, no entailment.

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

first, assum "not A"

4 clauses:

$$(A \lor B) \quad \underline{(\neg B \lor C)} \quad \underline{(\neg C)} \quad (\neg A)$$

resolve: $\neg B$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$$(A \lor B) \quad \underline{(\neg B \lor C)} \quad \underline{(\neg C)} \quad (\neg A)$$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $\quad$ $(\neg B \vee C)$ $\quad$ $(\neg C)$ $\quad$ $(\neg A)$ $\quad$ $(\neg B)$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$$(A \lor B) \quad (\neg B \lor C) \quad (\neg C) \quad (\neg A) \quad (\neg B)$$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

these Bs are complementary

$$(A \lor B) \quad (\neg B \lor C) \quad (\neg C) \quad (\neg A) \quad (\neg B)$$

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

add A

$\underline{(A \vee B)}$ $\quad (\neg B \vee C) \quad (\neg C) \quad (\neg A) \quad$ $\underline{(\neg B)}$ $\quad (A)$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$(A \lor B)$ $(\neg B \lor C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$ (A)

Can do all these iterations programatically
with iteration

Inference by Resolution

Does $(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C)$ entail A?

$$(A \vee B) \wedge (\neg B \vee C) \wedge (\neg C) \wedge (\neg A)$$

$(A \vee B)$ $\quad$ $(\neg B \vee C)$ $\quad$ $(\neg C)$ $\quad$ $\underline{(\neg A)}$ $\quad$ $(\neg B)$ $\quad$ $\underline{(A)}$

Inference by Resolution

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

when we resolve As

$$(A \lor B) \quad (\neg B \lor C) \quad (\neg C) \quad \underline{(\neg A)} \quad (\neg B) \quad \underline{(A)} \quad (\,)$$

= False ∴ contradic.

Inference by Resolution

and we know that A is true

Does $(A \lor B) \land (\neg B \lor C) \land (\neg C)$ entail A?

$$(A \lor B) \land (\neg B \lor C) \land (\neg C) \land (\neg A)$$

$(A \lor B)$ $(\neg B \lor C)$ $(\neg C)$ $(\neg A)$ $(\neg B)$ (A) $(\,)$

Resolution: keep looking at conclusions and see if we reach contradiction

Propositional logic is just 1 type and has limitations'

→ e.g. need a lot of symbols

First-Order Logic

↳ more powerful than propositional logic

Propositional Logic

Propositional Symbols

MinervaGryffindor

MinervaHufflepuff

MinervaRavenclaw

MinervaSlytherin

...

Every symbol here could either be T/F.

But relationships betw. symbols!

First-Order Logic → 2 types of symbols

e.g. objects / houses

<u>Constant Symbol</u>	<u>Predicate Symbol</u>
Minerva	*Person*
Pomona	*House*
Horace	*BelongsTo*
Gilderoy	
Gryffindor	
Hufflepuff	
Ravenclaw	
Slytherin	

relationships between symbols

↳ Properties that might hold true for constants.

First-Order Logic

Can use same symbols as in propositional

$Person(Minerva)$ Minerva is a person.

$House(Gryffindor)$ Gryffindor is a house.

$\neg House(Minerva)$ Minerva is not a house.

$BelongsTo(Minerva, Gryffindor)$

Minerva belongs to Gryffindor.

N.B. that only need 1 symbol per person and per house

2 main "quantifiers" in first-order logic:

1

Universal Quantification

lets us express that some statement will hold true for all values of x

Universal Quantification

"for all" → for all values of x, y holds true

$$\forall x.\ BelongsTo(x, Gryffindor) \rightarrow \neg BelongsTo(x, Hufflepuff)$$

For all objects x, if x belongs to Gryffindor, then x does not belong to Hufflepuff.

Anyone in Gryffindor is not in Hufflepuff.

2 **Existential Quantification**

says that some expression will be true for some values of some variable

Existential Quantification

there exists an x such that x is a house that Minerva belongs to

$$\exists x.\ House(x) \land BelongsTo(Minerva, x)$$

There exists an object x such that x is a house and Minerva belongs to x.

Minerva belongs to a house.

Quantifs. help us create more sophisticated statements than in propositional.

Existential Quantification

for all values of x if x is a person

$$\forall x.\ Person(x) \rightarrow (\exists y.\ House(y) \wedge BelongsTo(x, y))$$

For all objects x, if x is a person, then there exists an object y such that y is a house and x belongs to y.

Every person belongs to a house.

Knowledge

Introduction to

Artificial Intelligence

with Python