# Whodunnit? Crime Drama as a Case for Natural Language Understanding

**Lea Frermann**, Shay Cohen and Mirella Lapata

amazon

Edinburgh
University of Edinburgh
Natural Language Processing NLP

lfrerman@amazon.com

www.frermann.de

ACL, July 18, 2018

**Natural Language Understanding (NLU)**

- uncover information, understand facts and make inferences
- understand non-factual information, e.g., sentiment

# NLU as (visual) Question Answering

??

| | |
|---|---|
| In meteorology, **precipitation** is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include [...] | **Q:** *What causes precipitation to fall?* <br> **A:** *gravity.* |

?

**Q:** *Who is wearing glasses?*

**A:** *man.*

# NLU as Movie QA and Narrative QA

Movie QA from video segments (**?**)

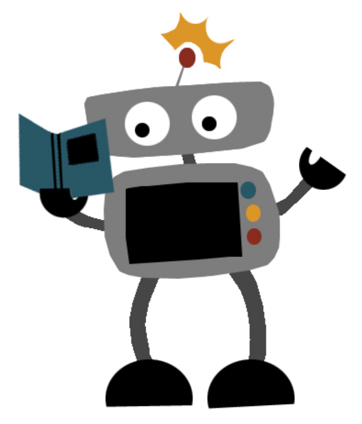
| | |
|---|---|
|  | **Q:** *Why does Forest undertake a 3-year marathon?*<br><br>**A:** *Because he is upset that Jenny left him.* |

Narrative QA from scripts and summaries (**?**)

| | |
|---|---|
| FRANK (*to the baby*) Hiya, Oscar. What do you say, slugger?<br>FRANK (*to Dana*) That's a good-looking kid you got there, Ms. Barrett. | **Q:** *How is Oscar related to Dana?*<br><br>**A:** *Her son* |

## NLU as Movie QA and Narrative QA

Movie QA from video segments (?)



**Q:** *Why does Forest undertake*

*a*

**A**

*t*

N... maries (?)

ya, Oscar.

s a good-

Ms. Bar-

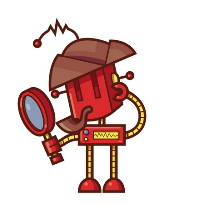
rett.

**A.** *Her son*

# This work: A new perspective!

**Tasks that are challenging for / interesting to humans**

- mysteries / questions with no (immediately) obvious answers
- non-localized answers
- accumulate relevant information



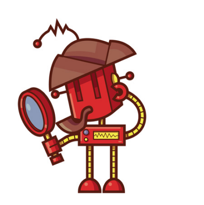
# This work: A new perspective!

**Tasks that are challenging for / interesting to humans**

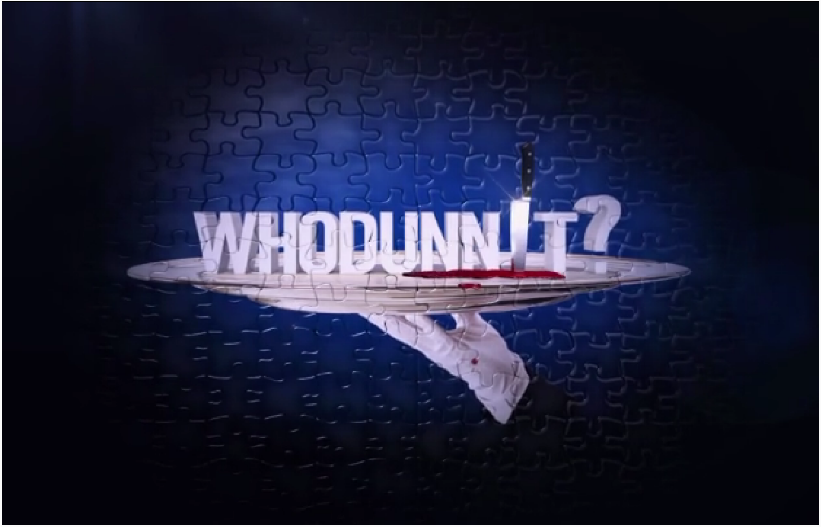
- mysteries / questions with no (immediately) obvious answers
- non-localized answers
- accumulate relevant information

**Towards Real-world Natural language inference**

- situated in time and space
- involves interactions / dialogue
- incremental
- multi-modal



# This work: A new perspective!



## This work: A new

# CSI as a dataset for real-world NLU


**Key Features**

- 15 seasons / 337 episodes → lots of data
- 40-64 minutes → manageable cast and story complexity
- schematic storyline
- clear and consistent target inference: **whodunnit?**

# The CSI Data Set

# Underlying Data (39 episodes)

1. DVDs → videos with subtitles

| | | |
|---|---|---|
| **Peter Berglund** | you 're still going to have to convince a jury that i killed two strangers for no reason | 00:38:44.934 |
| Grissom does n't look worried | | 00:38:48.581 |
| He takes his gloves off and puts them on the table | | 00:38:51.127 |
| **Grissom** | you ever been to the theater peter | 00:38:53.174 |
| **Grissom** | there 's a play called six degrees of separation | 00:38:55.414 |
| **Grissom** | it 's about how all the people in the world are connected to each other by no more than six people | 00:38:59.154 |
| **Grissom** | all it takes to connect you to the victims is one degree | 00:39:03.674 |
| Camera holds on Peter Berglund 's worried look | | 00:39:07.854 |

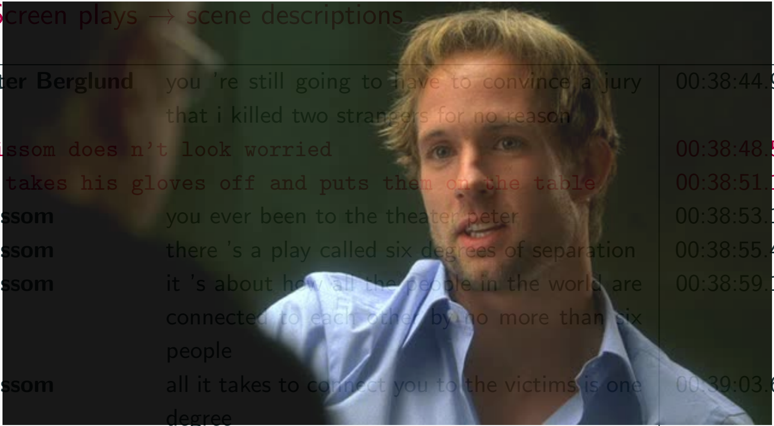
## Underlying Data (39 episodes)

1. DVDs → videos with subtitles
2. Screen plays → scene descriptions

| | | |
|---|---|---|
| **Peter Berglund** | you 're still going to have to convince a jury that i killed two strangers for no reason | 00:38:44.934 |
| `Grissom does n't look worried` | | 00:38:48.581 |
| `He takes his gloves off and puts them on the table` | | 00:38:51.127 |
| **Grissom** | you ever been to the theater peter | 00:38:53.174 |
| **Grissom** | there 's a play called six degrees of separation | 00:38:55.414 |
| **Grissom** | it 's about how all the people in the world are connected to each other by no more than six people | 00:38:59.154 |
| **Grissom** | all it takes to connect you to the victims is one degree | 00:39:03.674 |
| `Camera holds on Peter Berglund 's worried look` | | 00:39:07.854 |

# Underlying Data (39 episodes)

1. DVDs → videos with subtitles
2. Screen plays → scene descriptions



| | | |
|---|---|---|
| **Peter Berglund** | you 're still going to have to convince a jury that i killed two strangers for no reason | 00:38:44.934 |
| `Grissom does n't look worried` | | 00:38:48.581 |
| `He takes his gloves off and puts them on the table` | | 00:38:51.127 |
| **Grissom** | you ever been to the theater peter | 00:38:53.174 |
| **Grissom** | there 's a play called six degrees of separation | 00:38:55.414 |
| **Grissom** | it 's about how all the people in the world are connected to each other by no more than six people | 00:38:59.154 |
| **Grissom** | all it takes to connect you to the victims is one degree | 00:39:03.674 |
| `Camera holds on Peter Berglund 's worried look` | | 00:39:07.854 |

# Task Definition

# Whodunnit as a Machine Learning Task

## A multi-class classification problem

- classes $C = \{c_1, ..., c_N\}$ : $c_i$ participant in the plot
- incrementally infer distribution over classes

$$p(c_i = perpetrator | context)$$

☺ natural formulation from a human perspective

☹ strongly relies on accurate entity detection / coref resolution

☹ number of entities differs across episodes
$\rightarrow$ hard to measure performance

**A multi-class classification problem**

- classes $C = \{c_1, ..., c_N\}$ : $c_i$ participant in the plot
- incremental: infer distribution over classes

$$p(c_i = perpetrator|context)$$

☺ natural formulation from a human perspective

☹ strongly relies on accurate entity detection / coref resolution

☹ number of entities differs across episodes
$\leadsto$ hard to measure performance

# Whodunnit as a Machine Learning Task

**A sequence labeling problem**

- sequence $s = \{s_1, ..., s_N\}$ : $s_i$ sentence in the script
- incrementally predict for each sentence

$$\begin{cases} p(\ell^{s_i} = 1 | context), & \text{if perpetrator is mentioned in } s_i \\ p(\ell^{s_i} = 0 | context), & \text{otherwise} \end{cases}$$

☹ less natural setup from a human perspective

☺ incremental sequence prediction $\rightarrow$ natural ML problem

☺ independent of number of participants in the episode

# Annotation

# Annotation Interface

| Screenplay | Perpetrator mentioned? | Relates to case 1/2/none? |
|---|---|---|
| (Nick cuts the canopy around MONICA NEWMAN.) | | |
| **Nick** okay, Warrick, hit it | | |
| (WARRICK starts the crane support under the awning to remove the body and the canopy area that NICK cut.) | | |
| **Nick** white female, multiple bruising . . . bullet hole to the temple doesn't help | | |
| **Nick** .380 auto on the side | | |
| **Warrick** yeah, somebody manhandled her pretty good before they killed her | | |

# Annotation Interface

| Screenplay | Perpetrator mentioned? | Relates to case 1/2/none? |
|---|---|---|
| (Nick cuts the canopy around MONICA NEWMAN.) | ☐ | |
| **Nick** okay, Warrick, hit it | ☐ | |
| (WARRICK starts the crane support under the awning to remove the body and the canopy area that NICK cut.) | ☐ | |
| **Nick** white female, multiple bruising . . . bullet hole to the temple doesn't help | ☐ | |
| **Nick** .380 auto on the side | ☐ | |
| **Warrick** yeah, somebody manhandled her pretty good before they killed her | ☐ | |


1) Human guessing (IAA $\kappa = 0.74$)

# Annotation Interface

| Screenplay | Perpetrator mentioned? | Relates to case 1/2/none? |
|---|---|---|
| (Nick cuts the canopy around MONICA NEWMAN.) | ☐ | |
| **Nick** okay, Warrick, hit it | ☐ | |
| (WARRICK starts the crane support under the awning to remove the body and the canopy area that NICK cut.) | ☐ | |
| **Nick** white female, multiple bruising . . . bullet hole to the temple doesn't help | ☐ | |
| **Nick** .380 auto on the side | ☐ | |
| **Warrick** yeah, somebody manhandled her pretty good before they killed her | ☐ | |


1) Human guessing (IAA $\kappa = 0.74$)

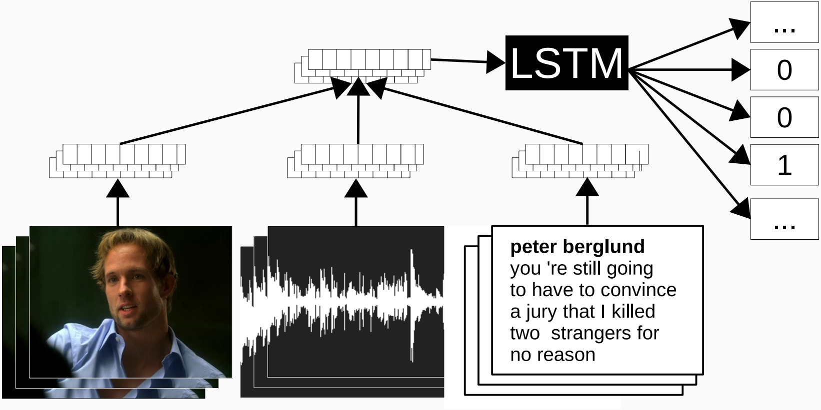
2) Gold standard (IAA $\kappa = 0.90$)

# An LSTM Detective

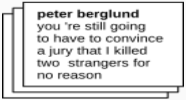
# Model: Overview

**Input** Sequence of (multi-modal) sentence representations

**Output** Sequence of binary labels:
perpetrator mentioned (1) / not mentioned (0)



# Input Modalities




sentence $s : \{w_1, ... w_{|s|}\}$

word embeddings, convolution and max-pooling

sound waves of video snippet of $s$

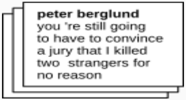
MFCCs for every 5ms

(background sound, music, no speech)

frame sequence of video snippet of $s$

sample one frame; embed through pre-trained image classifier (**?**)

# Input Modalities




sentence $s$ : $\{w_1, ...w_{|s|}\}$

word embeddings, convolution and max-pooling

sound waves of video snippet of $s$

MFCCs for every 5ms

(background sound, music, no speech)

frame sequence of video snippet of $s$

sample one frame; embed through pre-trained image classifier (?)

**Concatenate embedded modalities and pass through ReLu**

# Experiments

# Model Comparison

**Pronoun Baseline (PRO)**

- Simplest possible baseline
- predict $\ell = 1$ for any sentence containing a pronoun

# Model Comparison

**Pronoun Baseline (PRO)**

- Simplest possible baseline
- predict $\ell = 1$ for any sentence containing a pronoun

**Conditional Random Field (CRF)**

- Importance of sophisticated memory / nonlinear mappings
- graphical sequence labelling model

# Model Comparison

**Pronoun Baseline (PRO)**

- Simplest possible baseline
- predict $\ell = 1$ for any sentence containing a pronoun

**Conditional Random Field (CRF)**

- Importance of sophisticated memory / nonlinear mappings
- graphical sequence labelling model

**Multilayer Perceptron (MLP)**

- Importance of sequential information
- Two hidden layers and softmax output, rest like in LSTM

# Model Comparison

**Pronoun Baseline (PRO)**

- Simplest possible baseline
- predict $\ell = 1$ for any sentence containing a pronoun

**Conditional Random Field (CRF)**

- Importance of sophisticated memory / nonlinear mappings
- graphical sequence labelling model

**Multilayer Perceptron (MLP)**

- Importance of sequential information
- Two hidden layers and softmax output, rest like in LSTM

**Upper Bound (Humans)**

# Evaluation Metric

| speaker | utterance | perpetrator? gold | perpetrator? model |
|---|---|---|---|
| brass | mr heitz you 're mr newman 's realtor | 0 | 1 |
| augieheitz | what you kidding | 0 | 0 |
| augieheitz | my clients never have to see me | 0 | 0 |
| brass | you always give out the combination to your lockboxes | 0 | 0 |
| brass | it 's illegal | 0 | 1 |
| augieheitz | um you know i had a fish on the line | 0 | 0 |
| augieheitz | look | 0 | 0 |
| augieheitz | i only give out the combination to people that i really trust | 0 | 0 |
| brass nods his head as this makes perfect sense to him | | 0 | 0 |
| he looks over at grissom who does n't say anything | | 0 | 0 |
| catherine is interviewing peterberglund and the woman from the teaser | | 1 | 1 |
| she 's holding a bagged laptop in her arms | | 0 | 0 |
| catherine | all right look i read rooms for a living | 0 | 0 |
| catherine | that closet was tossed | 0 | 0 |
| catherine | the carpet lit up | 0 | 0 |
| catherine | so i 'm going to ask you again what were you doing in there | 1 | 1 |
| peterberglund | it was my idea | 1 | 0 |
| catherine | right | 0 | 0 |
| catherine | you did n't play with it too did you | 1 | 1 |
| nick is already at the edge of the pool | | 0 | 0 |
| he 's kneeling in front of something on the ground | | 0 | 0 |
| it looks like something reddish mixed with something else | | 0 | 0 |
| nick | hey warrick | 0 | 0 |
| warrick walks over to where nick is | | 0 | 0 |
| he also crouches down to look at what has nick 's attention | | 0 | 1 |
| warrick | yeah | 0 | 0 |
| nick | check this out | 0 | 0 |

...

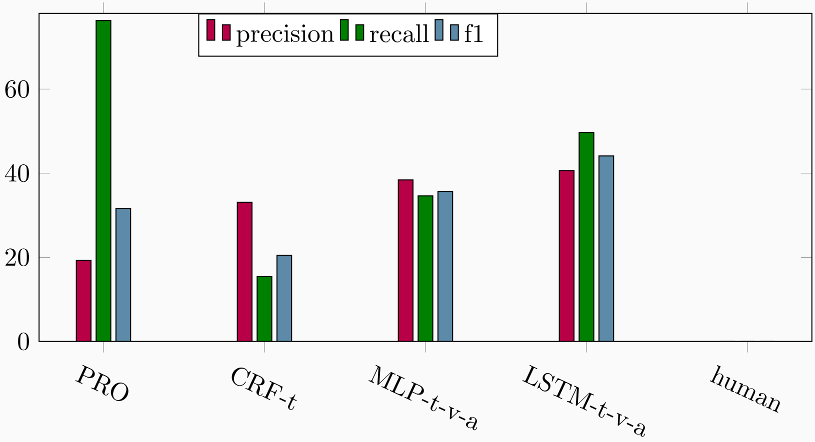
# Evaluation Metric

| speaker | utterance | perpetrator? gold | perpetrator? model |
|---|---|---|---|
| brass | mr heitz you 're mr newman 's realtor | 0 | 1 |
| augieheitz | what you kidding | 0 | 0 |
| augieheitz | my clients never have to see me | 0 | 0 |
| brass | you always give out the combination to your lockboxes | 0 | 0 |
| brass | it 's illegal | 0 | 1 |
| augieheitz | um you know i had a fish on the line | 0 | 0 |
| augieheitz | look | 0 | 0 |
| augieheitz | i only give out the combination to people that i really trust | 0 | 0 |
| brass nods his head as this makes perfect sense to him | | 0 | 0 |
| he looks over at grissom who does n't say anything | | 0 | 0 |
| catherine is interviewing peterberglund and the woman from the teaser | | 1 | 1 |
| she 's holding a bagged laptop in her arms | | 0 | 0 |
| catherine | all right look i read rooms for a living | 0 | 0 |
| catherine | that closet was tossed | 0 | 0 |
| catherine | the carpet lit up | 0 | 0 |
| catherine | so i 'm going to ask you again what were you doing in there | 1 | 1 |
| peterberglund | it was my idea | 1 | 0 |
| catherine | right | 0 | 0 |
| catherine | you did n't play with it too did you | 1 | 1 |
| nick is already at the edge of the pool | | 0 | 0 |
| he 's kneeling in front of something on the ground | | 0 | 0 |
| it looks like something reddish mixed with something else | | 0 | 0 |
| nick | hey warrick | 0 | 0 |
| warrick walks over to where nick is | | 0 | 0 |
| he also crouches down to look at what has nick 's attention | | 0 | 1 |
| warrick | yeah | 0 | 0 |
| nick | check this out | 0 | 0 |

...

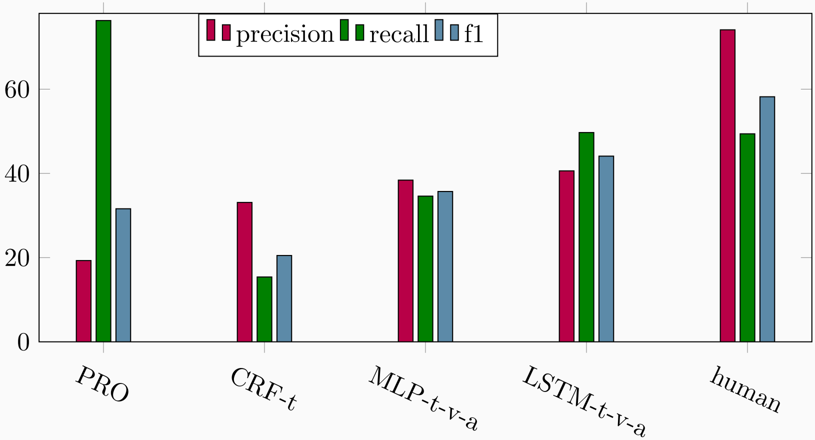
- **minority class**: perpetrator is mentioned ($\ell = 1$)
- precision / recall /f1

# Which Model is the Best Detective?



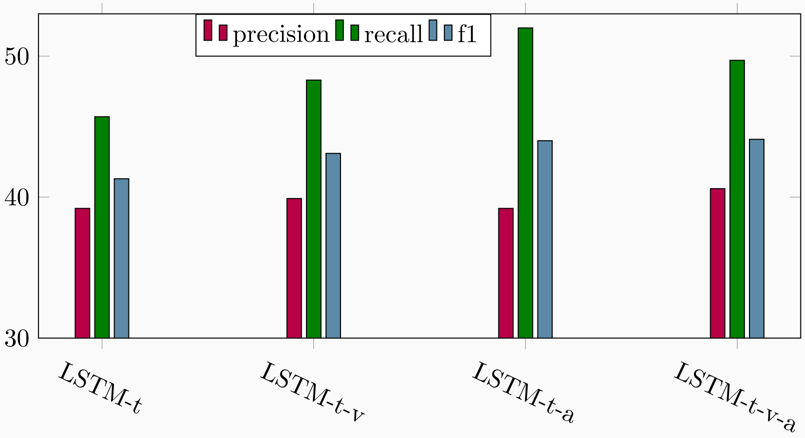

5-fold cross validation; 6 test episodes each

# Which Model is the Best Detective?




5-fold cross validation; 6 test episodes each

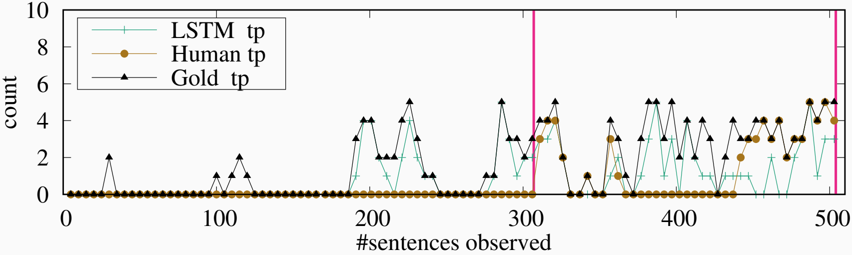
# Which Model is the Best Detective?




5-fold cross validation; 6 test episodes each

# Incremental Inference Patterns

Episode 19 (Season 03): "A Night at the Movies"



# Conclusions

# The end of police work as we know it?

# The end of know it?

8 Sun Saturday, November 11, 2017

THE SCOTTISH Sun SAYS

## Cold war

**NICOLA Sturgeon has had to put up with a lot from Alex Salmond since she took over as First Minister.**

Following ten years in his shadow, she stepped up to the big job which he had spent so long preparing her for.

There was no bitterness, no acrimony, no stab-in-the-back attack from a loyal protege. She simply took over the reins in a long arranged succession process.

Since then, Mr Salmond has on occasion made life difficult for the new First Minister with unguarded interventions on everything from Brexit to IndyRef2 and awkward Benny Hill-style gags.

So far Ms Sturgeon has done her best to laugh it off, treating his antics like those of an embarrassing elderly uncle.

She is conscious of all she owes him. He passed on everything he knew and she has benefitted hugely from his political wizardry, taking the party she inherited to even greater strengths.

There is a debt of gratitude and a burden of real friendship and admiration.

*But all the signs are that her patience is wearing thin. Her statement on Mr Salmond's decision to sign up with the Kremlin-funded RT TV station could hardly have been more chilly.*

She is stressing that she knew nothing of his contract. She is stressing that she was not consulted. And she is stressing that, given the chance, she would have advised against it.

Above all else, she is stressing Mr Salmond is a free agent, at liberty to work for anybody he wants. In short, he is nothing to do with her any more.

The statement from SNP HQ – and remember, Ms Sturgeon's husband is chief executive – was even more damning.

It pointed out a catalogue of human rights failings by the Russian government, from persecuting gays to invading neighbouring countries and insisted the SNP would never be quiet over that.

Mr Salmond's decision has split the SNP into two camps: loyal fans who think he can do no wrong and realists who see he has exposed the party and its indy ambitions to scorn and attack.

**The RT deal means Mr Salmond can be sure of fattening his bank account, but he has done it at the price of the respect and admiration of some of his closest friends.**

## Dark matter

**DEVOLVED governments aro… are "substantially in… Nicola Stur…**

1SM

# CRIME SCENE ARTIFICIAL INTELLIGENCE

# CSAI

## Tech that will change future of our policing

POLICE 2.0

**SCIENTISTS in Edinburgh this week revealed they had programmed computers that could solve cases from the hit US TV series CSI: Crime Scene.**

**Their computers managed to identify the murderer 60 per cent of the time – but did so much quicker than humans.**

**But Chief Features Writer MATT BENDORIS asks a retired top cop if this is one step closer to digital detectives.**

EX-detective Peter Ritchie believes it's only a matter of time before computers are solving crimes on their own – as they've already deleted the classic…

"Of course there's the romantic version of crime investigations as you see on TV or read in books. The reality is a lot different as there's also the impact of budgets, political implications, all these other factors that humans have to make outside the world of computers. But when it comes to analysing data, the computers can't be beaten."

Peter, from Port Seton, East Lothian, worked on the disappearance of five-year-…

…ton and Dinah McNicol. He says: "During my time in the police we had three serial killers on the go in Scotland at the same time with Black, Tobin and Sinclair.

"You can never say…

# The end of [illegible] know it

8 Sun Saturday, November 11, 2017

THE SCOTTISH Sun SAYS

Dark matter

THE SCOTSMAN Friday 10 November 2017

## AI computers learn to sleuth after watching TV crime show

By ILONA AMOS

Computers in a Scottish laboratory have been bingeing on box sets of a popular television crime drama in an effort to learn how to identify the culprit in each case.

The hit US show CSI, or Crime Scene Investigation, began in 2000 and ran for 15 seasons. Now it has been used for a novel experiment.

Scientists at the University of Edinburgh chose the series for a new study aimed at teaching machines how to solve a problem – in this case fingering a fictional killer – by assimilating information from images, audio, dialogue and scene descriptions.

They taught the artificially intelligent machines to approach solving the crimes in the same way people would – by considering which characters might be responsible on the basis of their behaviour in previous scenes.

Information in various forms – spoken, visual and textual – was processed as the plot of each episode unfolded.

They say the results of the study suggest such devices could play a role in developing efficient algorithms for real-world tasks that require complex reasoning.

Dr Lea Frermann, from the University of Edinburgh's School of Informatics, said: "Pinpointing the perpetrator in a TV show is a very difficult task for computers, but our model performed encouragingly well.

"We hope our findings will aid the development of machines that can take on board, and make sense of, large streams of information in real time."

The researchers set out to discover whether artificially intelligent computers can find the answers to puzzles that are challenging for humans.

They designed the computer model to solve arbitrary problems based on acquiring data.

The team mapped footage, script and background sounds from five seasons of the show into a machine-readable format and fed it into the computers, which learned to process the plot as each instalment unfolded and continually revised the criminal's identity.

The computers correctly identified the perpetrator during the final part of an episode 60 per cent of the time.

However, although the success rate for AI sleuths is impressive, it seems that they are not quite ready to replace flesh-and-blood detectives.

People who watched the same shows were able to work out who was responsible for the crimes 85 per cent of the time, the study found.

↑ Researchers used hit US TV show CSI to test artificial intelligence

COMMENT

"Pinpointing the perpetrator in a TV show is a very difficult task for computers, but our model performed encouragingly well"

DR LEA FRERMANN

SCOTSMAN.COM @THESCOTSMAN 17

## Scotland urged to ban bee-killing insecticides

By ILONA AMOS

The Scottish Government has come under fire from conservationists for "sitting on the fence" after Westminster said it would support a European ban on bee-killing pesticides.

Ministers have held off deciding on the future of neonicotinoid insecticide use in Scotland for too long and must now "get off the fence and show some leadership" to protect pollinators, according to the Scottish Wildlife Trust chief executive Jonny Hughes.

The UK government has announced it will back an EU proposal for an outright ban on the chemicals, but Scottish rural economy secretary Fergus Ewing said Scotland supports "continuation of the current restrictions", and will not decide until a review by the European Food Safety Authority is completed.

FARMING, PAGE 47

# Not quite...

**A general framework for incremental complex NLU**

- extensible e.g., with task-specific modules (entity disambiguation ...)
- generalizable across questions (‘where?’, ‘how?’, ...) and series

**(More) Faithful to human QA (in the wild)**

question $\rightarrow$ incrementally search ‘documents’ for the answer $\rightarrow$ stop once the answer is found

# Not quite...

## A new Task and Dataset

**Peter Berglund:**
You're still going to have to convince a jury that I killed two strangers for no reason.

*Grissom doesn't look worried.*
*He takes his gloves off and puts them on the table.*

**Grissom:**
You ever been to the theater Peter?
There 's a play called six degrees of separation.
It 's about how all the people in the world are connected to each other by no more than six people.
All it takes to connect you to the victims is one degree.

*Camera holds on Peter Berglund's worried look.*

| | | | | | |
|---|---|---|---|---|---|
| human predictions | 0 | 0 | 0 | 0 | 1 |
| gold standard | 1 | 0 | 1 | 1 | 1 |

https://github.com/EdinburghNLP/csi-corpus

# Not quite...

## A new Task and Dataset

**Peter Berglund:** You're still going to have to convince a jury that I killed two strangers for no reason.

*Grissom doesn't look worried. He takes his gloves off and puts them on the table.*

**Grissom:** You ever been to the theater Peter? There 's a play called six degrees of separation. It 's about how all the people in the world are connected to each other by no more than six people. All it takes to connect you to the victims is one degree.

*Camera holds on Peter Berglund's worried look.*

| | | | | | |
|---|---|---|---|---|---|
| human predictions | 0 | 0 | 0 | 0 | 1 |
| gold standard | 1 | 0 | 1 | 1 | 1 |

https://github.com/EdinburghNLP/csi-corpus

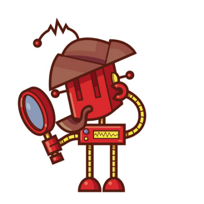
**Thank you!**

# Example LSTM Predictions

Episode 12 (Season 04): “Butterflied”

shots which truly mention the perpetrator


shots which the model predicts to mention the perpetrator



# Some Statistics on the CSI Dataset

| | |
|---|---|
| episodes with one case | 19 |
| episodes with two cases | 20 |
| total number of cases | 59 |

# Some Statistics on the CSI Dataset

| | | | | |
|---|---|---|---|---|
| | episodes with one case | 19 | | |
| | episodes with two cases | 20 | | |
| | total number of cases | 59 | | |
| | | min | max | avg |
| per case | sequence length (sents) | 228 | 1209 | 689 |
| | sentences with perpetrator | 0 | 267 | 89 |
| | scene descriptions | 64 | 538 | 245 |
| | spoken utterances | 144 | 778 | 444 |
| | characters | 8 | 38 | 20 |

# Some Statistics on the CSI Dataset

| | | | | | |
|---|---|---|---|---|---|
| | episodes with one case | | 19 | | |
| | episodes with two cases | | 20 | | |
| | total number of cases | | 59 | | |
| | | | min | max | avg |
| per case | sequence length (sents) | | 228 | 1209 | 689 |
| per case | sentences with perpetrator | | 0 | 267 | 89 |
| per case | scene descriptions | | 64 | 538 | 245 |
| per case | spoken utterances | | 144 | 778 | 444 |
| per case | characters | | 8 | 38 | 20 |
| | type of crime | murder | 51 | | |
| | type of crime | accident | 4 | | |
| | type of crime | suicide | 2 | | |
| | type of crime | other | 2 | | |

# Annotations: Summary

**1) Humans guessing the perpetrator** (IAA $\kappa = 0.74$)

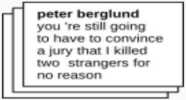
- binary sentence sentence-level tags
- real-time indications of humans (thinking they) know the perpetrator

**2) Gold standard** (IAA $\kappa = 0.90$)

- word-level indicators of {suspect, **perpetrator**, other} mentions
- This work: convert word-level tags to **sentence-level** labels

## Input Modalities: Text

**Raw text input** sentence $s : \{w_1, ...w_{|s|}\}$




- map words to pre-trained GloVe embeddings (50-dimensional)
- concatenate word embeddings
- pass vector through convolutional layer with max-pooling

# Input Modalities: Audio

**Raw audio input** sound waves of video snippet corresponding to sentence $s$

- all sound except spoken language (music, background, ...)
- extract Mel-frequency cepstral coefficients (MFCCs) for every five milliseconds
- 13-dimensional feature vectors
- sample and concatenate five vectors (equally spaced)

# Input Modalities: Video

**Raw visual input** frame sequence of video snippet corresponding to sentence $s$

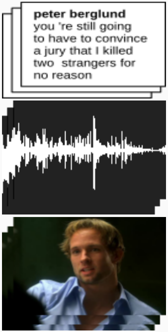
- sample one frame from the centre of the snippet
- pass through pre-trained CNN for object classification (inception-v4; **?**)
- use final hidden layer as visual feature vector

# Modality Fusion

**Modality fusion is learnt as part of the overall architecture**

- concatenate inputs
- pass through ReLu unit

$$xh = ReLU([\mathbf{x}^s; \mathbf{x}^a; \mathbf{x}^v]W^h + b^h)$$



## Settings

### Test Sets

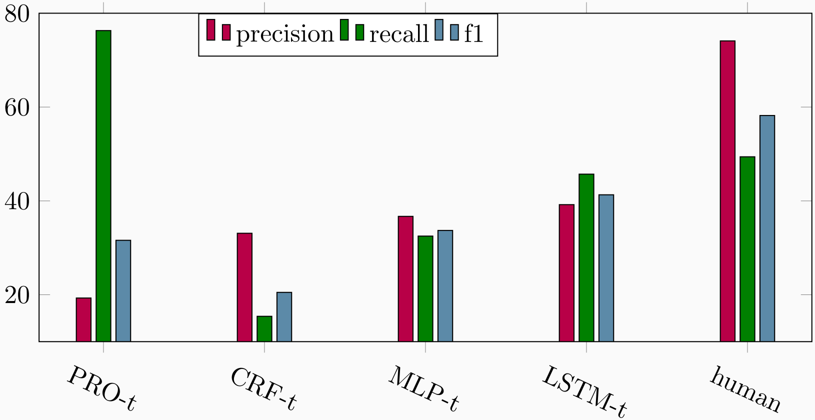
- 59 input sequences (each corresponding to one case)
- Cross-validation: 5 splits into 47 train / 6 test episodes
- Truly held-out set of 6 test episodes

### Training

- ADAM / SGD / Mini-batches
- Random initialization (except for word embeddings)
- Fine-tune word embeddings during training
- Train for 100 epochs; report best result
- Averages over five runs

# Which Model is the Best Detective?

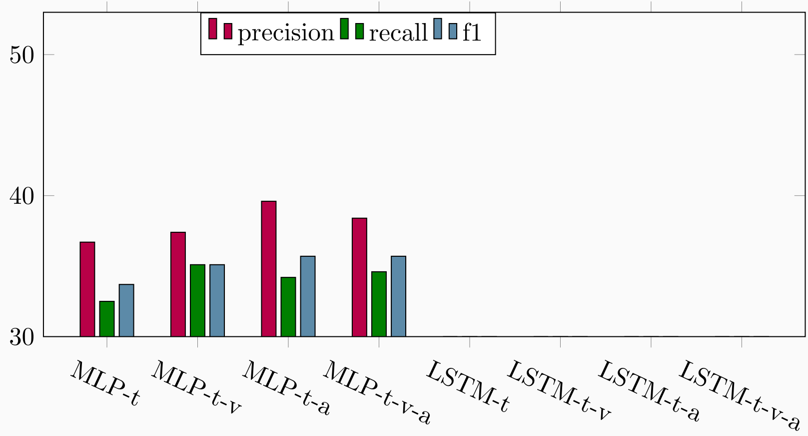
**All models: text only**




5-fold cross validation; 6 test episodes each

# Which Model is the Best Detective?

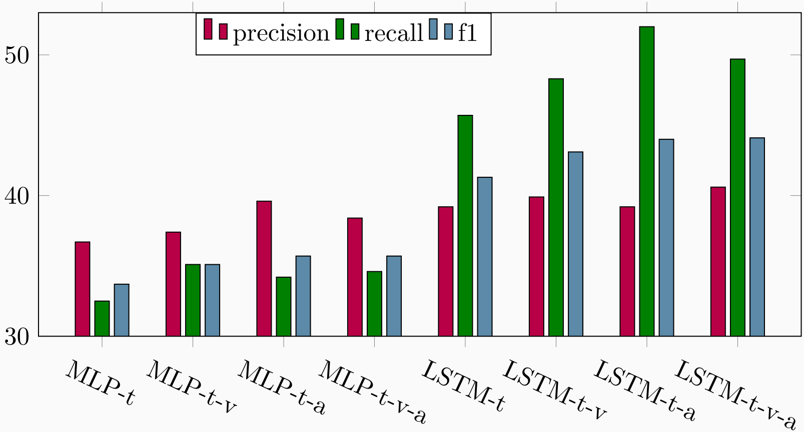
## MLP: all features




5-fold cross validation; 6 test episodes each

# Which Model is the Best Detective?

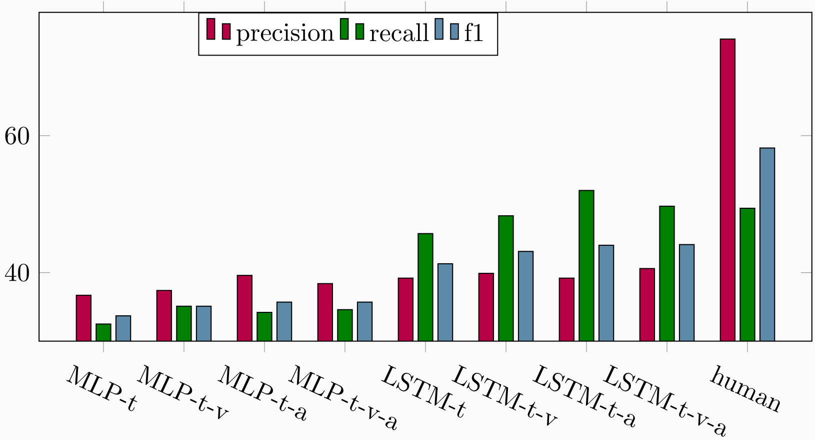
## + LSTM: all features




5-fold cross validation; 6 test episodes each

# Which Model is the Best Detective?

## + Humans




5-fold cross validation; 6 test episodes each

# Example LSTM Predictions

TODO CUT IF I DON'T HAVE TIME Episode 03 (Season 03): "Let the Seller Beware"

**saturation** → confidence that perpetrator is mentioned in sentence
**blue** → true perpetrator mentions

| s1 | s2 | s3 | s4 | s5 |
|---|---|---|---|---|
| *Grissom pulls out a small evidence bag with the filling* | *He puts it on the table* | *Tooth filling 0857* | *10-7-02* | **Brass** We also found **your** fingerprints and **your** hair |

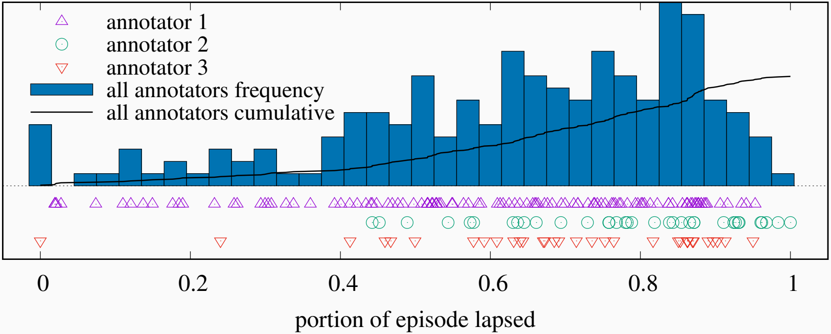
| s6 | s7 | s8 | s9 |
|---|---|---|---|
| **Peter B.** Look **I**'m sure you'll find **me** all over the house | **Peter B.** **I** wanted to buy it | **Peter B.** **I** was everywhere | **Brass** well **you** made sure **you** were everywhere too didn't **you**? |

# First correct perpetrator prediction

- At which point do humans / LSTM **correctly** predict the perpetrator **for the first time**?
- 30 test episodes used in cross-validation

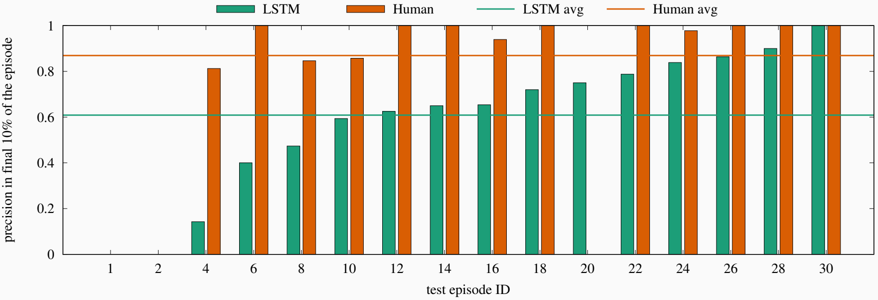
| | min | max | avg |
|---|---|---|---|
| LSTM | 2 | 554 | 141 |
| Human | 12 | 1014 | 423 |

# How do Humans Guess?



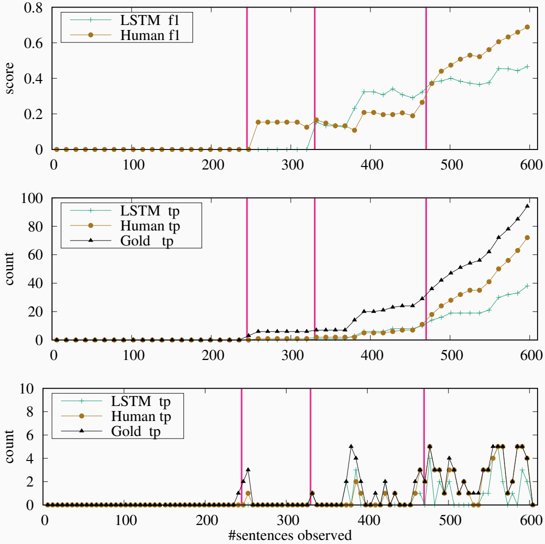
# Can the Model Identify the Perpetrator?

- In the last 10% of an episode: How precisely do humans / LSTM predict the perpetrator?
- 30 test episodes used in cross-validation



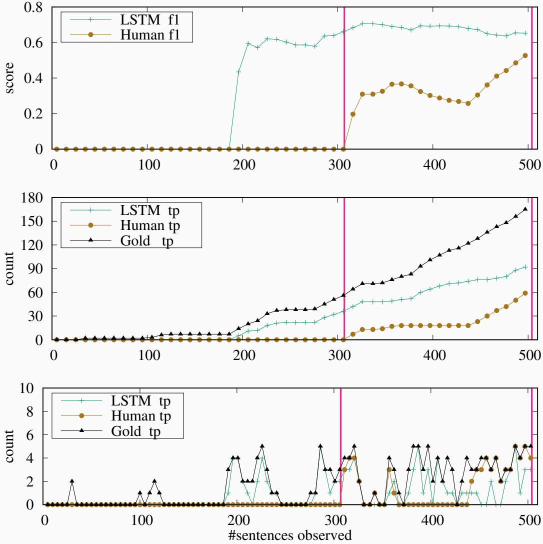
# Incremental Inference Patterns

Episode 12 (Season 03): "Got Murder?"



# Incremental Inference Patterns

Episode 19 (Season 03): "A Night at the Movies"



# What if there is no Perpetrator?

- LSTM (and humans!) are primed to expect a crime happening
- This case was a suicide
- Both humans and LSTM still predict a killer

# References

Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., and Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., and Blunsom, P. (2015). Teaching machines to read and comprehend. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in*

*Neural Information Processing Systems 28*, pages 1693–1701. Curran Associates, Inc.

Kociský, T., Schwarz, J., Blunsom, P., Dyer, C., Hermann, K. M., Melis, G., and Grefenstette, E. (2018). The NarrativeQA reading comprehension challenge. *Transactions of the Association for Computational Linguistics*, TBD:TBD.

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas, USA.

Szegedy, C., Ioffe, S., and Vanhoucke, V. (2016). Inception-v4, inception-ResNet and the impact of residual connections on learning. *CoRR*, abs/1602.07261.

Tapaswi, M., Zhu, Y., Stiefelhagen, R., Torralba, A., Urtasun, R., and Fidler, S. (2016). MovieQA: Understanding stories in movies through question-answering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4631–4640, Las Vegas, Nevada.