The Amazing Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives

Mohit Iyyer  Varun Manjunatha  Anupam Guha, Yogarshi Vyas, Jordan Boyd-Graber, Hal Daume III, Larry Davis

Presented by:
Leo Cao
Outlines

● Problem Definition
  ○ What is it & Why is it Hard
● Metrics (Loss Function)
● Data
● Methods
● Performance
The Problem is Hard

Comics = Stylized Artwork + Dialogue (Vision + Language)

- “Art” renders objects and concepts in many ways.
The Problem is Hard

Comics = Stylized Artwork + Dialogue (Vision + Language)

- Dialogues in comic often contain dialects and out-of-vocabulary tokens
- Judicious Omission

ACTION-TO-ACTION: 34.6%

NOW FOR THE BIG SURPRISE! YOU, A DETECTIVE, WILL DANGLE FROM THE NEXT POST!

HURRAY SO YOU'RE THE KILLER!

THROW THAT GUN AWAY!
Metrics that Make it Easier (Avoiding Text Generation)

Loss Function: Cross-Entropy

Three Types of Tasks: (Back to this later)

- Text Cloze
- Visual Cloze
- Character Coherence

Two Levels of Difficulty

- Easy: random candidates drawn from entire dataset
- Difficult: relevant candidates drawn from nearby sections

\[ s = \text{softmax}(A^T c), \]
Data

Golden Age Comics:

- Small self-contained stories (no backstories, no Batman)
- Public domain data
- Plenty of them

Processing:

- Panel Segmentation
- Text Segmentation
- OCR (Google Cloud Vision works the best, $3,000)
HAPPY LANDING, SISTER! HAW!!
OH POOR MR. HERBWEAVER!
HE'S GOING TO LOSE HIS JOB AND POOR ME! I'M GOING TO CRASH!

OH DEAR! WE'LL CRASH!!
YOU MEAN YOU'LL CRASH!! I'M GETTING OUT OF HERE!!

ALICE! I'VE BEEN LOOKING ALL OVER FOR YOU!

HIYA KID! ALL ALONE???

G-GOLLY! I'M STILL ALIVE! BUT WHERE AM I??
Thanks Old Timer! The bats wouldn't have got us, sure! Where'd they come from?

Scotty's my name. I'm the sheriff. Mean to tell you've never heard of the bats?

Scotty's my name. I'm the sheriff. Mean to tell you've never heard of the bats?

Thanks Old Timer! The bats wouldn't have got us, sure! Where'd they come from?

Back, Cora! A giant bat--and it moves so fast I can't seem to hit it!

Here comes another--and another! We can't escape!

Stop that, confused shootin', you two! I'm the law in this here town!

Hold that door open, mister!
## Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>easy</td>
<td>hard</td>
<td>easy</td>
</tr>
<tr>
<td>Random</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Text-only</td>
<td>62.6</td>
<td>51.9</td>
<td>57.5</td>
</tr>
<tr>
<td>Image-only</td>
<td>59.8</td>
<td>48.7</td>
<td>82.3</td>
</tr>
<tr>
<td>NC-image-text</td>
<td>63.1</td>
<td>59.6</td>
<td>-</td>
</tr>
<tr>
<td>Image-text</td>
<td><strong>71.0</strong></td>
<td><strong>63.4</strong></td>
<td><strong>83.0</strong></td>
</tr>
<tr>
<td>Human</td>
<td>–</td>
<td>84</td>
<td>–</td>
</tr>
</tbody>
</table>
Questions?