

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek^{1,2}, Josef Baloun^{1,2}, Martin Prantl^{1,2}, Ladislav Lenc^{1,2} & Pavel Král^{1,2}

¹Dept. of Computer Science & Engineering

²NTIS - New Technologies for the Information Society
University of West Bohemia
Plzeň, Czech Republic

<https://nlp.kiv.zcu.cz/>

May 2, 2024

Definition

- Dialogue Act (DA) = sentence level labels used for dialogue management and understanding as:
 - *Informs, Questions, Orders, Thanks, etc.*

Motivation & Goal

- DA recognition usually realized from a speech signal
- Little prior work on the printed dialogues
- → **Novel approach for DA recognition from image documents, namely comic books**
- New DA dataset

Approach Scheme

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

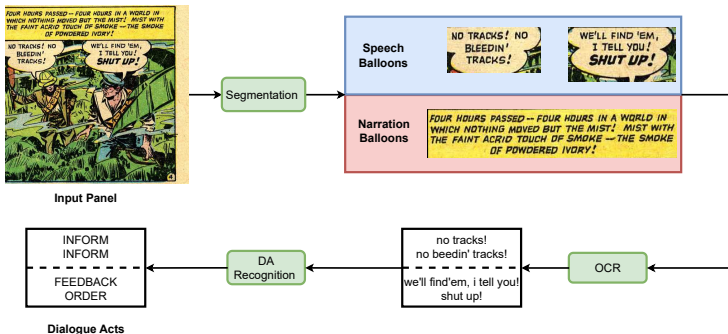


Figure: Pipeline of the proposed approach

Segmentation

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

- Regions of interest (RoI) detection and extraction
- Four object types
 - 1 Dialogues (speech balloons)
 - 2 Narration text (narration balloons)
 - 3 Onomatopoeia (comic drawings of words that phonetically imitate, resemble, or suggest the sound that they describe)
 - “Boom!” or “Bang!”
 - 4 Other text (e.g., shop names, bus stops, menu in restaurants, etc.)
- Two detection methods used and compared
 - YOLOv8 [1]
 - Faster R-CNN [2]

Optical Character Recognition

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

**Proposed
Approach**

Dataset

Experiments

Conclusions

References

- No general solution for OCR of comics
- **Google Vision OCR system**
 - Significantly better results on comics than other systems [3]
 - Trained on large corpora and gradually improved by Google company
- Balloon type classification
 - Transformer-based models as for the DA recognition

Transformer-based models

1 BERT model

- bert-base-uncased
- SwDA fine-tuned bert-base-uncased

2 RoBERTa model

3 Microsoft/DialoGPT model

- DialoGPT-small
- DialoGPT-medium
- DialoGPT-large

Dataset I

Basic Information

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

- Images mainly from COMICS public dataset¹
 - 800 annotated panels used from this dataset
 - 638 additional comic panels from other source
- Manual verification and error correction of segmentation and OCR errors

Dataset item	Counts
Panel	1438
Speech balloon	2196
Utterance (DA)	2282

Table: Main dataset information

¹<https://obj.umiacs.umd.edu/comics/index.html>

Dataset II

Additional information

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

- DAMSL annotation scheme [4]
- 15 DA classes
- Freely available for research purposes²



```
"dialog": [  
  "<BACKCHANNEL oh!> <BACKCHANNEL oh!>  
  <INFORM they must have radioed from the ship!>"  
],
```

Figure: Example of DA annotation

²<https://corpora.kiv.zcu.cz/comicorda/>

Dialogue Act Distribution

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

Dialogue act	Counts	%
Inform	1280	56.1
Backchannel	230	10.1
Order	192	8.4
Feedback	155	6.8
Wh_Question	107	4.7
Greet	73	3.2
Not_Classifiable	69	3.0
Yes_No_Question	64	2.8
Politeness_Formula	63	2.8
Commit	20	0.8
Thank	10	<0.5
Offer	9	<0.5
Or_Question	5	<0.5
Close	4	<0.5
Bye	1	<0.5

Table: DA distribution and frequencies in COMICORDA dataset

Comparison with other DA Datasets

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

SwDA		MRDA	
Statement-non-opinion	36%	Statement	59%
Acknowledgement	19%	Backchannel	14%
Statement-opinion	13%	Disruption	13%
Agreement	5%	Floorgrabber	7%
Abandoned	5%	Question	6%
Yes-No-Question	2%	-	-

VERBMOBIL		COMICORDA	
Feedback	26%	Inform	56%
Inform	20%	Backchannel	10%
Suggest	20%	Order	8%
Request	8%	Feedback	7%
Bye	4%	Wh_Question	5%
Greet	4%	Greet	3%

Experiments

Segmentation Results

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

Model	Speech Balloons	
	$AP^{iou=0.5}$	$AP^{iou=0.75}$
YOLOv8	0.954	0.823
Faster R-CNN	0.978	0.816

Table: Average precision results for speech balloons detection

Model	$mAP^{iou=0.5}$	$mAP^{iou=0.75}$
YOLOv8	0.896	0.739
Faster R-CNN	0.818	0.731

Table: Average segmentation results

- Better results with YOLOv8 model

Experiments

Narration vs Dialogue Text Classification

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

Model	GT text Accuracy	OCR Text Accuracy
SVM (baseline)	86.7	71.7
bert-base-uncased	93.4	93.7
RoBERTa-base	93.1	93.7
DialoGPT-small	92.1	92.0
DialoGPT-medium	93.4	93.2
DialoGPT-large	92.3	93.3

Table: Narration vs dialogue text classification results [in %]

- Comparable results on GT and OCR'd texts
- Comparable results performance across the models

Experiments

Dialogue Act Recognition

COMICORDA:
Dialogue Act
Recognition
in Comic
Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

Model	GT text		OCR text	
	Acc.	Macro-F1	Acc.	Macro-F1
SVM (baseline)	67.1	43.7	57.2	35.5
bert-base-uncased	77.7	55.5	67.5	48.3
SwDA bert-base-uncased	78.5	52.3	70.0	49.9
RoBERTa-base	75.9	51.9	65.5	43.0
DialoGPT-small	72.6	46.1	65.3	44.6
DialoGPT-medium	73.0	48.7	66.8	44.0
DialoGPT-large	77.1	54.1	67.7	45.2

Table: DA recognition results [in%]

- OCR errors → significant drop in DA accuracy
- BERT-based models - superior results

Conclusions

COMICORDA: Dialogue Act Recognition in Comic Books

Jiří Martínek,
Josef Baloun,
Martin
Prantl,
Ladislav Lenc
& Pavel Král

Introduction

Proposed
Approach

Dataset

Experiments

Conclusions

References

- First attempt for DA recognition from comic books
- Created a new COMICORDA dataset containing DA annotation from comic books, freely available for research
- Proposed an approach for DA recognition from image data
- Baseline results for future research in this area

References I

- [1] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with yolov8," *arXiv preprint arXiv:2305.09972*, 2023.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [3] M. Iyer *et al.*, "The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [4] J. Alexandersson *et al.*, "Dialogue acts in verbmobil-2," 1997.