How to write a great research paper
(according to Rob)

Until now

- Experimental setup
- Many interesting NLP tasks
- Language models
- Many interesting algorithms, architectures

Today

- Structure of research paper (in NLP)
- Suggestions per section
- General writing advice
- ChatGPT
- Exam
- Project
- Group formation

Today

~~You can only write a good paper if you have a talent for writing~~

Structure

What makes a great research paper?

- Addresses an important topic, task or issue (RQ)
- Advances our understanding
- Clearly written and easy to understand
- Provide an analysis to underline improvements
- Code available and re-usable

Anatomy of a Paper

Abstract

introduction Introduction

body

Related Work

Data

Analysis

Method

Results

conclusion Conclusions

References

Abstract

Introduction

Experiments

Results

Discussion

Related Work

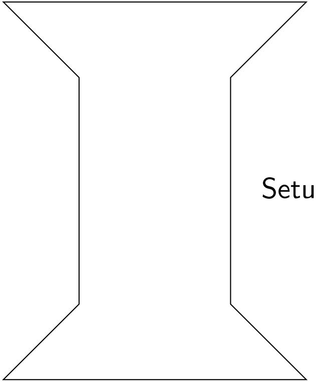
Conclusions

References

Section order and headings in the body may vary

Structure

Another view (the hourglass model):



Intro

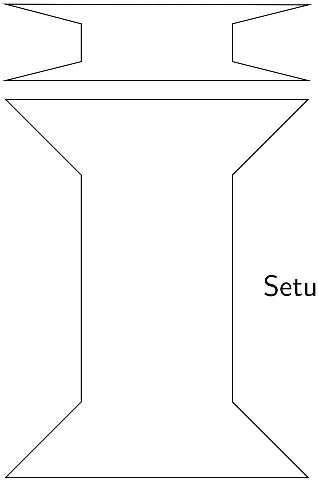
Rel. Work

Setup/Method/Results/Analysis

Discussion/Conclusion

Structure

Another view (the hourglass model):



Abstract

Intro

Rel. Work

Setup/Method/Results/Analysis

Discussion/Conclusion

Abstract

Note that it is a short version of your paper, a sales pitch!

- Summary of task and contributions
- Aimed at general audience
- Reader should be able to tell if your paper is relevant for their needs based on the abstract
- Write it last, after you've written and revised the whole paper
- Common to end with main finding, or bragging: our proposed model increases scores from X to Y on dataset Z.

Abstract

Because of globalization, it is becoming more and more common to use multiple languages in a single utterance, also called codeswitching.[1] This results in special linguistic structures and, therefore, poses many challenges for Natural Language Processing.[2] Existing models for language identification in code-switched data are all supervised, requiring annotated training data which is only available for a limited number of language pairs.[3] In this paper, we explore semi-supervised approaches, that exploit out-of-domain monolingual training data.[4] We experiment with word uni-grams, word n-grams, character ngrams, Viterbi Decoding, Latent Dirichlet Allocation, Support Vector Machine and Logistic Regression.[5] The Viterbi model was the best semi-supervised model, scoring a weighted F1 score of 92.23%, whereas a fully supervised state-of-the-art BERT-based model scored 98.43%.[6]

Taken from Iliescu et al. (2021): Much Gracias: Semi-supervised Code-switch Detection for Spanish-English: How far can we get?

Abstract

Because of globalization, it is becoming more and more common to use multiple languages in a single utterance, also called codeswitching.[1] This results in special linguistic structures and, therefore, poses many challenges for Natural Language Processing.[2] Existing models for language identification in code-switched data are all supervised, requiring annotated training data which is only available for a limited number of language pairs.[3] In this paper, we explore semi-supervised approaches, that exploit out-of-domain monolingual training data.[4] We experiment with word uni-grams, word n-grams, character ngrams, Viterbi Decoding, Latent Dirichlet Allocation, Support Vector Machine and Logistic Regression.[5] The Viterbi model was the best semi-supervised model, scoring a weighted F1 score of 92.23%, whereas a fully supervised state-of-the-art BERT-based model scored 98.43%.[6]

1: General topic

2: Problem

3: Current state

4: Solution/proposed direction

5: Details proposed approach

6: Brag about performance (conclusion)

Introduction

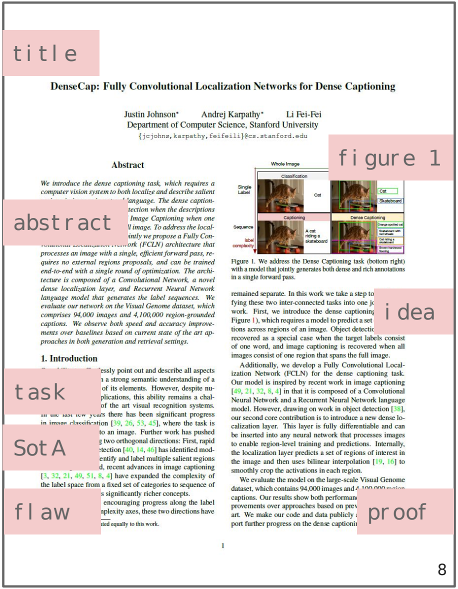
Often considered to be the hardest to write!

- Task and its importance
- State-of-the-art, standard practice, or common assumption
- Flaw in state-of-the-art, standard practice, or common assumption
- Your idea/solution: contributions/research questions
- (Proof it works?)

Introduction

One example says more than a thousand words!

Introduction



DenseCap: Fully Convolutional Localization Networks for Dense Captioning

Justin Johnson* Andrej Karpathy* Li Fei-Fei
Department of Computer Science, Stanford University
{jcjohns,karpathy,feifeili}@cs.stanford.edu

Abstract

We introduce the dense captioning task, which requires a computer vision system to both localize and describe salient … language. The dense caption- … tection when the descriptions … Image Captioning when one … ll image. To address the local- … intly we propose a Fully Con- … ork (FCLN) architecture that processes an image with a single, efficient forward pass, requires no external regions proposals, and can be trained end-to-end with a single round of optimization. The architecture is composed of a Convolutional Network, a novel dense localization layer, and Recurrent Neural Network language model that generates the label sequences. We evaluate our network on the Visual Genome dataset, which comprises 94,000 images and 4,100,000 region-grounded captions. We observe both speed and accuracy improvements over baselines based on current state of the art approaches in both generation and retrieval settings.

1. Introduction

… essly point out and describe all aspects … n a strong semantic understanding of a … of its elements. However, despite nu- … plications, this ability remains a chal- … of the art visual recognition systems. … there has been significant progress in image classification [39, 26, 53, 45], where the task is … to an image. Further work has pushed … g two orthogonal directions: First, rapid … tection [40, 14, 46] has identified mod- … entify and label multiple salient regions … d, recent advances in image captioning [3, 32, 21, 49, 51, 8, 4] have expanded the complexity of the label space from a fixed set of categories to sequence of … s significantly richer concepts.

… encouraging progress along the label … nplexity axes, these two directions have

… ted equally to this work.

Whole Image | Classification | Cat | Cat | Skateboard | Single Label | Sequence | label complexity | Captioning | A cat riding a skateboard | Dense Captioning | Orange spotted cat | Skateboard with red wheels | Cat riding a skateboard | Brown hardwood flooring

Figure 1. We address the Dense Captioning task (bottom right) with a model that jointly generates both dense and rich annotations in a single forward pass.

remained separate. In this work we take a step to … fying these two inter-connected tasks into one jo … work. First, we introduce the dense captioning … Figure 1), which requires a model to predict a set … tions across regions of an image. Object detectio … recovered as a special case when the target labels consist of one word, and image captioning is recovered when all images consist of one region that spans the full image.

Additionally, we develop a Fully Convolutional Localization Network (FCLN) for the dense captioning task. Our model is inspired by recent work in image captioning [49, 21, 32, 8, 4] in that it is composed of a Convolutional Neural Network and a Recurrent Neural Network language model. However, drawing on work in object detection [38], our second core contribution is to introduce a new dense localization layer. This layer is fully differentiable and can be inserted into any neural network that processes images to enable region-level training and predictions. Internally, the localization layer predicts a set of regions of interest in the image and then uses bilinear interpolation [19, 16] to smoothly crop the activations in each region.

We evaluate the model on the large-scale Visual Genome dataset, which contains 94,000 images and … captions. Our results show both performan … provements over approaches based on prev … art. We make our code and data publicly … port further progress on the dense captioni …

1

Source and more detailed guide:
`https://twitter.com/kate_saenko_/status/1371884306470219779`

Literature review/Related work

DO:

- Present previous works that:
  - address the same issue
  - attempt to solve the same task
  - use similar research methods
- Briefly describe their methods and findings
- Explain how they are related to your research
- Point out a limitation or gap
- Briefly explain how your work addresses these limitations or gaps

DO NOT:

- Recount the entire history of the research problem
- Diminish previous work

Literature review/Related work

DO:

- Present previous works that:
  - address the same issue
  - attempt to solve the same task
  - use similar research methods
- Briefly describe their methods and findings
- **Explain how they are related to your research**
- Point out a limitation or gap
- Briefly explain how your work addresses these limitations or gaps

DO NOT:

- Recount the entire history of the research problem
- Diminish previous work

Literature review/Related work

How to structure related work?

- Depends on how they relate, try to find commonalities
- Could be 1 paragraph per paper
- But in many cases it's better to group them!

Literature review/Related work

Do not do this at the end!

- You might be reinventing many wheels

Methodology

DO:

- Describe your whole pipeline
- Include following sections (if you have these parts):
  - Data description (if a contribution$\mapsto$separate section)
  - Data preparation and pre-processing
  - Feature engineering
  - Model architecture
  - Model training
- Pay special attention to your own contributions
- Provide justification for each decision
- Make sure your evaluation metric(s) are appropriate for the task

DO NOT:

- Include an extensive theoretical background of your methods
  - For example a whole section on how BERT works is irrelevant in many cases (except if you're improving most parts)
  - This is what citations are for

Experiments

Combine?:

- Data
- Setup
- Results

Experiments

Combine?:

- Data
- Setup
- Results

Decide based on your contributions!

Experiments

Data:

- if new: motivate (probably in own section, with collection decisions, annotation decisions etc.)
- else: describe, and motivate why this one was chosen

Experiments

Setup:

- Mostly necessary if non-standard setup is used.
- Could also be used to include metrics, language model selection, and even datasets

Results

2 options:

- Results without interpretation (have it in discussion)
- Include interpretation in results

Results

Always compare to baseline!:

- The simplest possible approach (majority baseline, i.e. everything is positive or noun)
- A simple machine learning classifier (logistic regression with words as features)
- The "state-of-the-art" approach on which you want to improve (your starting point)

Results

Figures and tables

- Make sure every piece of information can be interpreted properly:
  - which dataset/split?
  - which model?
  - which metric?
- This include clear (standalone) captions, and clear column/row y-axis/x-axis titles
- Guide the reader through the interesting results/findings

Analysis

- Not always mentioned in standard research paper structure
- Highly relevant for NLP!

Analysis

What to include

- Quantitative: observe statistics
  - For inspiration for NER, see: https://aclanthology.org/2020.emnlp-main.489/
- Qualitative: look at data

Analysis

Quantitative analysis:

- Performance per class (i.e. confusion matrix)
- Other metrics (i.e. precision/recall)
- Ablation/Isolation testing of parts of the model

Analysis

Qualitative analysis:

- Look at instances:
  - Are there trends in mistakes of model A
  - Where/why does model A do better on certain instances compared to model B
- Could provide interesting example sentences in paper, or even summarize quantitatively

Discussion

DO:

- Summarize key findings
- Interpret the results and try to answer your research question(s)
- Discuss wider implications and/or prospective applications
- Consider alternative explanation(s) of your findings
- Acknowledge the limitations
- Make recommendations for further research

DO NOT:

- Introduce new methods or results
- Make inflated claims
- Diminish your research

Future work

Personally, I'm not a big fan. I would suggest to discuss limitations instead (Note: subjective)

- Especially in the research paper of this course, there is normally no future work.
- But also for research papers; they are a product by themselves, not a series.

Conclusion

- Restate your task and why it is important
- Restate your claim(s)
- Summarize your methods and findings
- Address opposing viewpoints and/or shortcomi

Bibliography

- parenthetical: \cite{Knuth1997}
  X can be defined as Y (Knuth, 1997).
- narrative: \newcite{Knuth1997}
  Knuth (1997) defines X as Y.
- We're in luck: `https://aclanthology.org/` is an amazing resource
- **use it!**

General writing advice

- Follow the style guide (citations)
- Audience; your peers
- **How much detail: everything relevant for RQ**
- Appendices: should be optional for reader
- Use active phrases for things you did

General writing advice

Separate what you want to tell in a tree structure

- The (sub)sections
- But also the paragraphs
- You can use latex comments for this

General writing advice

Separate what you want to tell in a tree structure

```
\subsection{current state}
% Introduce standard splits

% A test set can only be used N times

% What was the situation before
```

General writing advice

Separate what you want to tell in a tree structure

```
\subsection{Current State}

% Introduce standard splits
\label{sec:splits}
In Natural Language Processing (NLP), a highly empirical field, it is common to
benchmark multiple models to each other on a standard dataset. However, since
most current models are supervised, and thus require labeled training data, the
datasets have to be split. To ensure a fair comparison, most datasets in NLP
have standard splits. Most datasets consist of three splits (also visualized in
Figure~\ref{fig:splits}(a)):
\begin{itemize}
    \item \textbf{train}: Used for training models, in some setups this split
can be omitted (zero-shot or unsupervised learning).

    \item \textbf{dev}elopment (also called validation/evalua- tion): Used to c
different versions of the proposed model(s). Can also be used to get preliminary
to the main research questions.

    \item \textbf{test}: Used to confirm the final answer to the research quest
\end{itemize}

\begin{figure}
    \centering
    \input{imgs/splits}
    \caption{Overview of the use of data splits. \colorbox{red}{red}:test
\colorbox{orange}{orange}:dev \colorbox{green}{green}:train
\colorbox{yellow}{yellow}:tune. a) standard splits for traditional machine
learning models b) standard splits as used for neural network models c) our
proposed splits for neural network models.}
    \label{fig:splits}
\end{figure}

% A test set can only be used N times
One often raised worry is that if too many papers are written based on the same
test-set, overfitting occurs, especially when only positive results are
published~\cite{scargle2000publication}. It should be noted that we do not
refer to overfitting of the models parameters, but on design decisions
(hyperparameters etc.), in line with ``bias from research design'' as defined
by~\newcite{hovy2021five}. This means that there is a bias towards methods that
perform well on this specific set. We agree that this is a danger. If we
consider a more general perspective to this problem, a certain split becomes
more prone to this when more different models are evaluated on this exact same
data. Let's assume that there is a threshold $N$ that limits the number of
times we can re-use the same split for evaluation. The number of papers that
can use the same dataset for a fair comparison is then equal to $N$ divided by
the average number of evaluated models per paper. From this, it follows that,
no matter how large $N$ is, a larger average number of runs per paper will
drastically reduce the lifespan of a dataset.
```

General writing advice

Why not to cite blogs/arxiv papers:

- https://robvanderg.github.io/blog/nlp.htm
- Arxiv practicaly only checks whether your latex code is valid

General writing advice

Common pittfalls:

1. Do not assemble your paper as a patchwork of your sources
   Readers want your work and analysis, not a summary of your sources
2. Do not organize your paper as a narrative of your thinking
   Readers don't want to know what you found first, or all paths you explored
3. Keep the experimental standards and bias lecture in mind:
   - Do significance testing, random bootstrap for single runs, ASO for multiple seeds:
     `https://github.com/Kaleidophon/deep-significance`
4. Use the provided style files and correct citations (ACL Anthology)

General writing advice

When to write a research paper?

- Start early
- Trying to explain your ideas uncovers missing motivations
- Details are still fresh

General writing advice

tricks:

- If uncertain where to put results, put them in analysis section (larger analysis, looks like you went more in-depth)
- Put "boring" details in appendix
- The grade will be decided by your teacher, it might make sense to take a look at 1 or 2 of their (or their students) papers
- Analysis is what distinguishes a great from a good paper
- Think about your title

General writing advice

Do organize your paper around the core elements of your argument: your claim and the reasons supporting it
Questions?

ChatGPT

We follow:
`https://2023.aclweb.org/blog/ACL-2023-policy/`

- Add a section in appendix if you made use of a chatbot (since we do not use a Responsible NLP Checklist)
- Include each stage on the ACL policy, and indicate to what extend you used a chatbot
- Use with care!, you are responsible for the project and plagiarism, correctness etc.

Exam

4 persons

- 10 minutes presentation
- 5 minutes clarification questions for group
- 10-12 minutes per person: questions project, and random topic (and experimental standards/bias)
- 10-15 minutes grading

Exam

You can expect a variety of types of questions, like:

- Walk us through algorithm X
- How does method X differ from method Y
- What are the benefits of using method X for task Y

Exam

Suggestions presentation:

- 10 minutes is short; overview is often not even necessary
- It's like the paper: focus on your contributions
  - What do you want to convince me of?
  - Ask for each piece of content, is this relevant?

Project phase

- Hand in baseline predictions on LearnIt: 27-03
- Project proposal presentations: 29-03 and 03-04
- Project proposal deadline: 12-04
- Paper draft: 19-05
- Paper: 26-05