# Multi-documentation Summarization of Science Articles

Justin To, Luka Liu, Milan Dean

W 266: Natural language Processing UC Berkeley School of Information

April 14, 2023

#### Abstract

In today's age of information overload, the ability to generate accurate and concise summaries of large bodies of text is more important than ever. This study focuses on the task of abstractive multi-document summarization (MDS), which involves generating a summary based on multiple related documents. Two state-of-the-art pre-trained large language models (LLMs), Centrum and LED, were selected, and different strategies, including fine-tuning and model-stacking, were explored for building an MDS model customized for Multi-XScience, a dataset previously unseen by the pre-trained models. In evaluating the summaries generated, we adopted both quantitative ROUGE-based scoring and qualitative analysis of the model outputs, in order to understand how the models adapted to unique elements specific to the task and dataset. By studying the process and strategies for adapting pre-trained models to domain-specific MDS tasks, we aim to contribute to the research on adapting pre-trained models for new and domain specific MDS tasks and datasets.

# 1 Introduction

MDS is the task of summarizing multiple texts into a concise and informative summary. Compared to single document summarization, it is challenging in the context of abstractive MDS to keep the summary coherent and comprehensive, given the documents are longer<sup>1</sup> and of a more complex structure<sup>2</sup>. Despite the challenges, MDS has a potentially wider application as real-world tasks often involve collating and summarizing information from multiple sources. An important question therefore is, given the advances in NLP and the availability of pre-trained LLMs, how should we build a customized abstractive MDS model if we are given a domain-specific dataset previously unseen by the models?

In this paper, we employ the Multi-XScience dataset, a relatively recent dataset that has not been widely used for pre-training/fine-tuning LLMs, to simulate the situation where we are given an MDS task for a broad domain, e.g. academic writing from the sciences. The defining characteristics of the dataset, namely (a) the significant portion of samples with long inputs (up to  $\sim 6300$  tokens) and (b) that different articles within each sample are related vet not covering the exact same event<sup>3</sup>, led us to pick two models for experimentation. The first is the Longformer Encoder Decoder (LED) model, which was proposed by Beltagy et al. [1] and introduced the concept of local vs global attention so that attention-based transformers can still effectively handle the computation of long inputs. The second is the Centrum model, proposed by Puduppully and Steedman [6], for its capability of handling long inputs and inclusion of a centroid-based approach for dealing with multiple source documents.

Leveraging on the available pre-trained check-

points, we first tested the off-the-shelf LED and Centrum using the Multi-XScience dataset and then proceeded to fine-tune them for improved performance. Then, we explored the use of a two-step setup whereby individual source articles of each sample are first shortened through summarization, and then combined again for MDS For evaluation, we employed the processing. common ROUGE score metric, showing the improved performance from fine-tuning (ROUGE-2 from 5.2 to 6.9; ROUGE-L from 14.6 to 17.8, which are close to state-of-the-art (SOTA)). Further analyses were conducted on the variation of performance across inputs of different lengths, and the amount of "copying" from the main article. Lastly, a qualitative analysis was done on 25 samples for each model tested, and the fine-tuned and two-step models were found to demonstrate capabilities in extracting information from and contrasting different sources, while adapting its writing style to one suited for the dataset.

# 2 Related Works

In recent years, there has been significant research in the field of MDS, with a particular focus on the use of transformer-based models for abstractive models and through approaches such as sparse attention mechanisms, hierarchical sentence representations, and document-level clustering. Much progress has also been made in providing datasets for MDS training and evaluation.

On model development, the LED model proposed by Beltagy et al. [1] addresses the issue of processing long documents by introducing a sparse attention mechanism that allows the model to scale to documents of up to 16384 tokens. This approach has outperformed previous models, including advanced ones such as PEGASUS, on long-document tasks such as summarization.

Branching off to models more specialized in MDS, PRIMERA, proposed by Xiao et al. citexiao2022primera, is a pre-training method for transformer-based models that leverages hierarchical sentence representations to improve per-

<sup>&</sup>lt;sup>1</sup>For instance, the widely used CNN/Daily Mail dataset has an average input length of around 780 to-kens. In contrast, around 30% of the inputs in the Multi-XScience dataset has over 1024 tokens, while the longest sample has over 6300 tokens.

<sup>&</sup>lt;sup>2</sup>In MDS, the source articles are written by multiple authors with varying writing styles and article structures.

<sup>&</sup>lt;sup>3</sup>Further discussed in Section 3.1

formance on MDS tasks. The approach involves constructing a sentence-level pyramid structure and applying a masked language modeling objective to the pyramid, leading to improved results over BART, PEGASUS, and LED.

Advancing the research on MDS further, Puduppully and Steedman [6] proposed a centroid-based pre-training method for MDS that leverages document-level clustering to capture document-level semantics. This approach showed improved performance on summarization tasks over previous models including PRIMERA, particularly for MDS datasets with a large number of documents.

On the dataset front, a commonly used dataset in the MDS domain is perhaps the Multi-news dataset, which was introduced by Fabbri et al. [2] and contains news articles from multiple sources. Other MDS datasets include Wikisum, which was generated by Liu et al. [4] from Wikipedia articles, WCEP compiled by Ghalandari at el. [3] based on news summaries from the Wikipedia Current Events Portal, and the Multi-XScience dataset by Lu et al. [5] used in this study.

### 3 Methods

#### 3.1 Dataset

We used the Multi-XScience MDS dataset [5] in this study, which is a collection of scientific articles from scientific fields, including computer science, physics, and biology. Some key parameters of the dataset are set out as follows, while Appendix A summarizes other findings and visualizations from our EDA, as well as the data preprocessing steps adopted for the experiments.

The dataset is chosen for two reasons. First, Multi-XScience is among the few datasets previously unseen by MDS models during pretraining<sup>4</sup>. Picking datasets already seen by the models could pose difficulties in evaluation, not to mention that it would go against the overall

	Training	Validation	Test
Sample size	30369	5066	5093
Input Token Length	899 565 82-4694	898 567 79-4183	886 547 131-6348
Samples with inputs > 1024	21 21 04	20 71 87	20.070
tokens	31.21%	30.71%	30.87%
Label Token Length	$   \begin{array}{r}     142 \\     58 \\     24-735   \end{array} $	$     \begin{array}{r}     141 \\     58 \\     25-324     \end{array} $	$ \begin{array}{r} 142 \\ 58 \\ 26-418 \end{array} $
Articles per sample	4.43 2.62 2-21	4.43 $2.63$ $2-20$	$\begin{array}{c} 4.39 \\ 2.55 \\ 2-19 \end{array}$

Table 1: Dataset description. The cells with three columns refer to mean/standard deviation/range

goal to explore MDS model-building strategies for unseen data.

The second reason is that Multi-XScience is a more challenging task that is more comparable to real-world data collation and summarization. Many of the datasets (e.g. Multi-news, WCEP) are news-based, with the articles within a sample covering the same news event and key elements to be included in the summary appear-The Multi-XScience ing in multiple sources. dataset however, expects the model to write the related works section of a journal paper based on (a) the abstract of that paper which serves to provide context; and (b) the abstracts of the other journal articles that the main paper referenced. While these additional abstracts are usually<sup>5</sup> closely related, they often concern different aspects of an issue, and the model needs to be able to compare and contrast the similarities and differences<sup>6</sup>. This difficulty of the dataset also shows up empirically in [8].

 $<sup>^4\</sup>mathrm{LED}$  's pre-training data included Multi-News while Centrum's pre-training included both Multi-News and WikiSum.

<sup>&</sup>lt;sup>5</sup>Some exceptions exist, e.g. sample 845 (refer to Appendix D for sample review for sample review)) of the test set includes an irrelevant article on the design attributes of luggage carriers when the main article is on a MDS over online product reviews. See Appendix C for details.

<sup>&</sup>lt;sup>6</sup>In contrast, the commonly used Multi-news dataset has, in each sample, multiple news reports surrounding a single event, and the overall theme to be covered in the summary is generally present in every source article.

<sup>&</sup>lt;sup>7</sup>For instance, the zero- and few-short evaluation of BART, PEGASUS, LED and PRIMERA vielded a

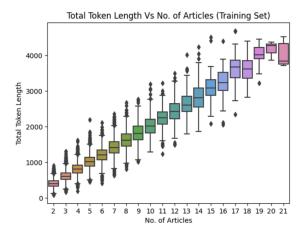


Figure 1: Total token length as a function of the number of articles in the training data set. The relationship between token length and the number of articles is linear.

As can be seen from the table above, the input length of the Multi X-Science dataset shows great variability, with some inputs so short that they can fit in more traditional summarization models, while others exceed even 4096, which is the max input length for Centrum. The same variability can also be observed in terms of the number of articles per sample (see Figure 1 below), with the two distributions showing a close-to-linear pattern. This provides an opportunity for us to use the Multi X-Science dataset to explore possible differences in model performance between short and long samples.

#### 3.2 Models

Due to resource constraints, we focused on two pre-trained LLMs, namely LED and Centrum for our experiments. Of the three recent and well-performing models we reviewed, we did not pick PRIMERA for two reasons - (a) Centrum is, based on literature review, the best-performing model while LED is a common starting point for both PRIMERA and Centrum, meaning it could be treated as a more advanced baseline;

ROUGE-2 score of only 1.9 to 4.6 and a ROUGE-L score of 9.9 to 15.7 for Multi-XScience, while the score ranges for Multi-news is 3.7 to 13.6 for ROUGE-2 and 10.4 to 20.8 for ROUGE-L. A similar gap remains even after fine-tuning of the models.

and (b) the authors of PRIMERA already provided performance results of PRIMERA on the Multi-XScience test set.

We adopted two different baselines for performance comparison. The first one, dubbed the "baseline model" is simply a lead-based model, whereby the first 3 sentences are chosen as the summary and is a common primary strategy implemented in various papers (e.g. [5], [7] and [9]). This also provides a useful baseline for comparing the degree of "copying" (see Section 3.3) from the main article which is an undesirable trait for the chosen dataset. The second baseline is the publicly available checkpoint for LED, which serves as the base architecture for recent MDS LLMs.

On top of the 2 baseline models, we ran experiments on the performance of the following models, with details on their fine-tuning and inference settings provided in Appendix B:

- 1. an alternative checkpoint of LED fine-tuned on the arXiv dataset
- 2. publicly available checkpoint of Centrum;
- 3. a version of the baseline LED<sup>8</sup> that we finetuned on the Multi-XScience training set;
- 4. a similarly fine-tuned version of Centrum; and
- 5. a two-step model stacking the fine-tuned LED and Centrum.

Of these experiments, models (c) and (d) stem from our plan to test out the performance of fine-tuning strategies for MDS. During the process, we noticed that despite the larger input length of Centrum (4096 tokens), there are still cases where the inputs have to be truncated and information omitted. We therefore explored under model (e) a two-step model stacking process whereby the source articles are first condensed

<sup>&</sup>lt;sup>8</sup>In theory, it would be better to fine-tune the alternative LED checkpoint which (a) is a larger version of LED with double the parameters; and (b) has been fine-tuned on text from the scientific fields. However, the GPU we had was unable to train the larger model while maintaining the max input token length at a reasonable level.

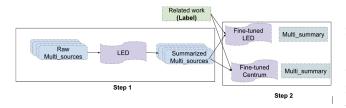


Figure 2: Two-steps model design

by our fine-tuned LED individually followed by an MDS process by (a) the fine-tuned LED or (b) the fine-tuned Centrum. The idea is to investigate if Centrum performs better when presented with more concise sources and without token truncation, see Figure 2. Refer to Appendix B for more details on the two step model.

#### 3.3 Evaluation metrics

As in the standard practice for summarization, the primary metric we use is the ROUGE score. Among the different versions of ROUGE, we chose ROUGE-2 and ROUGE-L, with a particular focus on the latter given the prevalence of compound nouns in academic writings<sup>9</sup>, for gauging model performance. Compared to ROUGE-2, ROUGE-L is better suited for this task. In calculating these metrics, precision, recall, and f-measure are all considered<sup>10</sup>, and scores were computed for the overall test dataset, as well as the subsets of short, medium, and long token length samples<sup>11</sup>.

Unlike other MDS datasets where the key information is expected to be found in most, if not all, of the source articles, the Multi X-Science dataset has a different structure whereby the main article (i.e. first source) is expected to provide general context while the other sources will each provide a related, yet different angle of the topic. It is desirable for models to extract information not just from the main article, and we therefore measured the degree of "copying" from the main article by repurposing standard ROUGE calculations<sup>12</sup>.

Finally, we supplemented our evaluation with a qualitative analysis whereby 25 random samples were selected and we reviewed the summaries written by the baseline LED and the 5 models experimented, judging them based on fluency, ability to extract information from multiple sources, and contrast information from the main article and other sources, as well as factual accuracy.

# 4 Results and Discussion

## 4.1 Overall performance

In the following table, we report the ROUGE-2 (R-2) and ROUGE-L (R-L) scores of the 2 baselines plus 5 models we tested, with precision (P), recall (R), and f-measure (F) scores presented. Scores for overall, short, medium, and long samples are also presented.

As shown in the tables above, the fine-tuned models (LED and Centrum) provided a markedly improved performance over the baselines and off-the-shelf models. In fact, even though the models were only fine-tuned for up to 2 epochs<sup>13</sup>, our figures quickly approached the SOTA figures reported in [8] (i.e. ROUGE-2 of 6.8; ROUGE-L of 18.2) demonstrating the viability of LED and Centrum to be adopted for previously unseen MDS data. When considering the breakdowns, the improvement for longer samples was much

<sup>&</sup>lt;sup>9</sup>This is an empirical observation from viewing samples in the dataset, and is quite intuitive considering academic terms are often ones glued together from multiple words (e.g. the theme of "Synchronous Optical Network Ring Assignment Problem" in sample 4820 of the test set, and "fixed-size ordinally forgetting encoding" in sample 3157).

<sup>&</sup>lt;sup>10</sup>We also look only at the mean score but not the lower and upper ranges due to the limited scope of our study, though we recognize that looking at the variation in performance across samples is in itself a valid future investigation direction.

<sup>&</sup>lt;sup>11</sup>We divided the 5093 test samples into three groups. "Short" samples consisted of 1273 samples with less than 486 tokens (lower quartile) for the inputs; "medium" with 2546 samples with input lengths between 486 and 735 (upper quartile) tokens; and the remaining 1274 samples with input lengths greater than 1150 tokens are grouped into the "long" samples.

<sup>&</sup>lt;sup>12</sup>Implemented using the standard rouge function by replacing the target reference that the function takes in with the main article of the sample.

 $<sup>^{13}</sup>$ Training details and considerations are at Appendix B.

more visible, indicating the MDS-specific model architectures are perhaps more suitable for tasks with longer input lengths.

This result is consistent with what our qualitative analysis (Appendix C). For example, in sample 485, the Centrum model was able to extract the key ideas of 3 relevant studies and contrast them with the main article, forming a coherent summary, while the baseline and off-the-shelf models were overwhelmed by the large number of articles (9 in total) and choose only to copy one of the articles in full. This is very different from shorter samples, e.g. sample 4160 with only 3 articles, where the off-the-shelf models were still able to prepare a summary covering elements, albeit in a more extractive manner without efforts to change the tone or merge the extracted parts together.

For the two-step model, the ROUGE scores do not show any advantage of the approach over simple fine-tuned models. However, as we observe from the qualitative analysis (e.g. samples 831 and 5068), there is a strong tendency for the two-step model to highlight differences and similarities between the main and other articles. However, the downside is that the two-step model made quite a number of factual fallacies in the process<sup>14</sup>.

# 4.2 Degree of Copying

To assess the degree of copying from the main article, we calculated a re-purposed ROUGE-L score. As shown in the table below, the fine-tuned and two-step models exhibited a significantly lower degree of "copying", which is supported by the qualitative analysis showing that these models were able to extract information from multiple articles. In contrast, the baseline LED model showed the highest degree of "copying", often copying substantial portions of the main article to the point that the number of

Model	R-2	Overall	Short	Medium	Long
D 1:	P	4.50	4.78	4.68	3.87
Baseline	R	8.13	6.12	7.87	10.7
(Lead)	F	5.40	5.02	5.56	5.48
Baseline	Р	4.52	4.12	4.75	4.47
(LED)	R	6.99	8.17	7.07	5.63
(LED)	F	5.21	5.2	5.40	4.83
	Р	5.25	5.32	5.30	5.09
LED	R	6.31	6.97	6.34	5.57
	F	5.39	5.58	5.44	5.11
	P	4.14	4.02	4.23	4.09
Centrum	R	8.12	9.48	8.14	6.70
	F	5.24	5.36	5.34	4.93
	P	8.75	8.03	8.67	9.63
Our LED	R	5.89	5.21	5.69	6.94
	F	6.56	5.83	6.41	7.62
Our	P	7.62	7.10	7.58	8.28
Centrum	R	7.27	6.38	7.13	8.44
	F	6.88	6.14	6.78	7.82
2-Step	Р	5.87	5.66	5.92	5.99
(Centrum)	R	6.67	6.15	6.62	7.26
(Centrum)	F	5.86	5.42	5.87	6.28

Table 2: Rouge-2 Scores for Models Tested (Best in **bold**; worst is *italicized*)

Model	R-2	Overall	Short	Medium	Long
Baseline	Р	12.4	14.9	12.7	9.17
_ 0.00 0 0	R	21.7	18.8	21.3	25.2
(Lead)	F	14.6	15.5	15	12.9
Baseline	Р	12.9	11.6	13.3	13.5
(LED)	R	20.5	23.8	20.4	17.3
(LED)	F	14.9	14.8	15.1	14.6
	Р	15.1	15.4	15.1	14.3
LED	R	18.6	20.3	18.7	16.7
	F	15.7	16.2	15.7	15.0
	Р	11.3	10.8	11.4	11.7
Centrum	R	22.7	25.5	22.8	19.7
	F	4.4	14.4	14.6	14.2
	P	23.6	24.2	23.5	23.2
Our LED	R	15.9	15.7	15.5	16.8
	F	17.7	17.6	17.4	18.4
Our	Р	19.8	20.9	19.6	19.1
Centrum	R	18.9	18.4	18.7	19.7
Centrum	F	17.8	17.9	17.6	18.1
2-Step	Р	17.0	18.0	17.0	15.9
(Centrum)	R	19.6	19.7	19.5	19.6
(Centrulli)	F	17.0	17.4	17.0	16.7

Table 3: Rouge-L Scores for Models Tested (Best in **bold**; worst is *italicized*)

<sup>&</sup>lt;sup>14</sup>One possible reason is that we need better tuning on the amount and manner of information fed into the 2nd step model, but resource limitations were a main obstacle considering the generation of the first step answers alone took over 14 hours.

Model	R-2	Overall	Short	Medium	Long
Baseline	Р	39.7	56.6	39.7	22.7
(Lead)	R	40.7	41.1	39.0	43.9
(Leau)	F	36.8	44.9	36.8	28.5
Baseline	Р	76.9	83.1	86.9	50.7
(LED)	R	75.5	92.8	81.5	46.3
(LED)	F	74.2	85.6	81.9	47.2
	Р	45.3	58.8	43.9	34.7
LED	R	33.6	44.3	31.5	27.0
	F	37.4	48.9	35.6	29.5
	Р	48.2	56.5	48.6	39.2
Centrum	R	57.5	73.7	55.7	45.1
	F	51.2	62.4	50.8	41.0
	Р	33.6	35.5	34.7	29.5
Our LED	R	13.5	13.0	13.2	14.7
	F	18.3	18.0	18.2	18.7
Our	Р	29.4	32.9	29.4	26.1
Centrum	R	17.3	16.4	16.9	19.0
Centrum	F	20.5	20.7	20.1	20.9
2-Step	Р	24.7	27.9	25.0	21.0
(Centrum)	R	17.2	17.1	16.8	18.0
(Centrulli)	F	19.1	20.1	18.9	18.6

Table 4: Rouge-L Scores for Measuring "Copying" (Best in **bold**; worst is *italicized*)

copied sentences exceeded even that of the leadbased baseline.

Although the ROUGE scores did not reveal any significant "copying" for the off-the-shelf LED and Centrum models, it is noteworthy that these models occasionally copied from a single reference article rather than the main article (e.g. see the output of the base Centrum for sample 4371). This indicates that there is still room for improving the "copying" metric, such as by taking the minimum of the ROUGE-L scores compared to each input article (rather than just the main article) to discourage "copying" from a single source.

#### 4.3 Other observations

The qualitative analysis also allowed us to make the following observations on the models, including:

1. The fine-tuned and two-step models showed clear signs of learning common writing styles (using phrases such as "there is a large body

- of work...") which is desirable in real world applications;
- 2. For samples containing irrelevant articles, the fine-tuned models were able to safely navigate away from them (e.g. sample 845)
- 3. Despite having been fine-tuned on journal articles, the models showed greatest difficulties when presented with highly technical papers that use common words in uncommon meanings (e.g. in mathematics in sample 4858). This shows the potential to enhance performance by adding another fine-tuning step on domain-specific texts. Alternatively, the two-step model has continued to perform well in such cases which shows the potential of such an approach.

#### 5 Conclusion

In conclusion, we demonstrated the viability of building custom-based MDS models through fine-tuning which is able to quickly approach SOTA scores even with simple and limited tuning. Through breakdown and qualitative analyses of the results, we highlighted the importance of considering the token length of input texts when selecting a text summarization model. Our results suggest that the fine-tuned LED and Centrum models are much more adept at MDS for longer texts, while the two-step model shows potential especially for highly technical texts despite needing more tuning. Future work could focus on further optimizing the tuning techniques (e.g. fine-tuning with domain-specific texts first before the MDS data), improving on the two-step approach, as well as looking at the performance stability of the models across data samples.

# **Appendix**

# A EDA and Data Preprocessing

This appendix provides further findings from the EDA on the Multi-XScience dataset, as well as the data preprocessing procedures undertaken in the experiments.

As mentioned above, one standard summary could be summarized from more than 20 source articles. With concatenating all sources of articles, the total token length is about 4096. In addition, we've examined the token length of these standard summaries (label) and provided the details below.

To determine the input length, we look at the average length of the documents we want to summarize. With the consideration of the various numbers of source documents and our GPU capabilities, we decided to take input numbers of tokens = 512 as an experiment and numbers of tokens = 4096 (max tokens in the training dataset) as our final model.

For the output length, considering the desired level of detail in the summary and the distribution of standard summaries (label) tokens, we decided a summary length between 100-250 may be sufficient.

Multi-XScience includes a main abstract, which is from the main article; a standard summary of the related works with a separation of "@cite", which is the summary of multi-source documentations, label datasets; and reference abstracts, which are the multiple abstracts from a different source of works. We processed the dataset by defining the separator of documents as "|||||", a document separator, then concatenating reference abstracts using the document separator to replace citation references.

#### B Model

#### B.1 Baseline Model

Copying the first 3 sentences from each reference abstract was used as our baseline. It is a simple and easy-to-implement approach that provides a starting point for summarizing multiple documents. It assumes that the first few sentences of a document contain important information that should be included in the summary.

This approach can serve as a useful baseline for several reasons:

- 1. It is easy to implement and does not require a lot of computational resources. This makes it a good starting point for developing more advanced summarization models.
- 2. It is straightforward to understand and explain. This makes it a good choice for initial experiments and evaluations.
- 3. It can provide a quick estimate of the performance of a summarization system. By comparing the summaries generated by this approach to human-written summaries or other machine-generated summaries, it is possible to get a rough idea of how well the system is performing.

# B.2 Baseline Base LED (1K and 16K) Model

The base LED model has ("allenai/led-base-16384"), with max input token length of 16384 & ~200M parameters and consists of 12 transformer encoder layers and 2 transformer decoder layers. This model is designed to be computationally efficient and suitable for low-resource environments. The baseline LED was generated based on a pre-trained LED model by Beltagy, L (2020) [1] on X-science test datasets with input tokens varying from 1k and 6k.

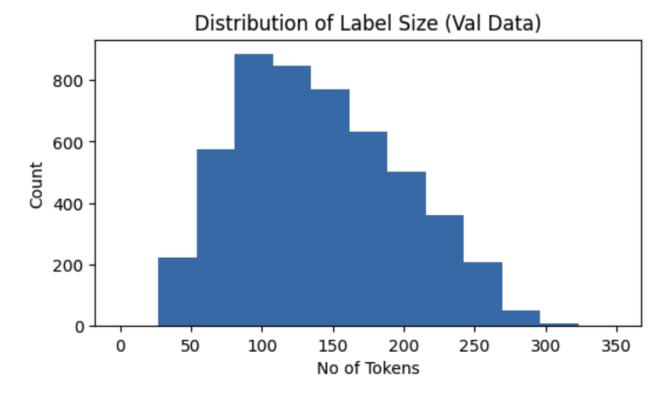


Figure 3: Distribution of the number of tokens in the label of training data. The number of tokens is right-skewed and centered around 100.

# B.3 Baseline Large LED (1K and 16K) Model

The "large" in the model name indicates that it has more parameters and is larger than the "base" LED model. The model has ("allenai/led-large-16384-arxiv"), with max input token length of  $16384 \& \sim 512 \mathrm{M}$  parameters, and consists of 24 transformer encoder layers and 4 transformer decoder layers. This model has been pre-trained for general summarization tasks. The baseline LED was generated based on a pre-trained large LED and evaluated on X-science test datasets.

#### B.4 Baseline Centrum Model

The "ratishsp/Centrum" model with max input token length of  $4096 \& \sim 192 \mathrm{M}$  parameters and consists of 12 transformer encoder layers and 12 transformer decoder layers. It is smaller than some of the larger transformer models, such as T5 and GPT-3, but larger than some of the smaller transformer models, such as the BART

base. The publicly available Centrum checkpoint was built upon the LED model architecture using the 4096 token version. This means Centrum is not able to take in 16384 tokens.

#### B.5 Fine-tuned LED Model

Fine-tuning the "LED-large-16384-arxiv" model to the X-science train (size 30369) and validation (size 5066) datasets specifically for the related work summarization. We used 4096 max number of tokens in tracking and validation sets, and 256 max output tokens with a batch size of 2 (due to limited GPU resources). The model is then used as a cross-entropy loss function for 2 epochs, and evaluated on test datasets(size 5093) with ROUGH metrics.

#### B.6 Fine-tuned Centrum Model

Fine-tuning the "ratishsp/Centrum" model to the X-science train and validation datasets specifically for the related works summarization. We used 4096 max number of tokens in tracking and validation sets, and 256 max output tokens with a batch size of 16 (due to limited GPU resources). The model is then used as a crossentropy loss function for 2 epochs, and evaluated on test datasets with ROUGH metrics.

### B.7 Two-step LED Model

In the first step, the LED model is used to generate a summary of each related source document. By summarizing each source document, the model can condense the information and reduce redundancy, making it easier to process and analyze. For the first step, none of the individual source articles in the X-science dataset contain more than 4096 tokens, so there is no issue of the first step model (i.e. the fine-tuned LED) receiving truncated inputs.

In the second step, generate the summary based on the fine-tuned LED model to refine the summaries generated in the first step. In this way, we were hoping that the fine-tuned LED model can learn to generate even more accurate and relevant summaries.

# B.8 Two-step Centrum Model

The same first step as the two-step LED Model above. In the second step, generate the summary based on the fine-tuned Centrum model to refine the summaries generated in the first step. In this way, we were hoping that the fine-tuned Centrum model can learn to generate even more accurate and relevant summaries.

# B.9 Two-step Model results

However, from our evaluation and observation of this approach, the performance of the two-step model does not exceed the fine-tuned LED/Centrum model. One potential issue is that the large amount of information contained in the source documents may make it difficult for the model to generate concise and informative summaries. Ultimately, the performance is heavily dependent on the performance of step 1.

Overall, our results demonstrate the importance of utilizing advanced techniques like finetuning to unlock the full potential of language models in real-world applications.

# C Qualitative Analysis

We also investigated how the model performed when summarizing short, medium, and long documents. The study aimed to evaluate the quality of the generated summaries and identify any trends or patterns in the model's performance based on the length of the source documents. A total of 25 randomly drawn samples (from the 5093 samples in the X-Science test dataset) are analyzed. These 25 samples include:

- 10 short samples, i.e. with total token lengths below the lower quartile
- 10 medium samples, i.e. with total token lengths between the lower and upper quartiles
- 5 long samples, i.e. with token lengths above the upper quartile

Our analysis of the review (refer to the Analysis Appendix for more detailed review) results showed that the fine-tuned LED and Centrum learned the structure for multi-documents summary for X-science dataset, with particularly good results for longer documents (ref. Sample 485) while for shorter samples the advantage over the baseline and off-the-shelf models appear to be less as those models are able to extract information from multiple sources to some degree, albeit with limited ability to merge them into a single coherent summary (ref. Sample 4160).

Additionally, we observed that LED and Centrum seem to have difficulties in digesting math topics, where the use of common language for unusual meanings might have confused the models (refer to no 4858). The two-step model somehow helps for these samples, but the downside is that the two-step model hallucinates quite a bit.

In general, LED and Centrum never really hallucinate. To future improve this issue, we would suggest that further research in this area should consider:

- Add domain-specific knowledge: This can include adding specialized math dictionaries or knowledge bases to your training data.
- Use special tokens for math symbols: This can help the model recognize and differentiate between mathematical expressions and regular language. For example, use a special token such as "[math]" to indicate the start of a math expression and "[/math]" to indicate the end.
- Use more training data: Consider adding more math-specific data to the training set to help the model learn to recognize and understand math concepts.
- Fine-tune your model: fine-tuning it on a smaller dataset of math-specific documents.
   This can help the model learn to recognize and summarize math-related information more effectively.

Overall, our study highlights the importance of considering the length of source documents, and the training datasets topics when evaluating the performance of multi-document summarization models.

Future research in this area should focus on developing models that can effectively summarize long documents; and provide the model with more specialized knowledge to help the model recognize and differentiate between regular language and math expressions, improving its ability to summarize math-related documents effectively. Refer to the manual review documentation in more detail link.

# D Detailed Review Samples

Refer to the following pages.

### References

- [1] Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer, 2020.
- [2] Alexander Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir Radev. Multinews: A large-scale multi-document summarization dataset and abstractive hierarchical model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1074–1084, Florence, Italy, July 2019. Association for Computational Linguistics.
- [3] Demian Gholipour Ghalandari, Chris Hokamp, Nghia The Pham, John Glover, and Georgiana Ifrim. A large-scale multi-document summarization dataset from the wikipedia current events portal, 2020.
- [4] Peter J. Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences, 2018.
- [5] Yao Lu, Yue Dong, and Laurent Charlin. Multi-XScience: A large-scale dataset for extreme multi-document summarization of scientific articles. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8068–8074, Online, November 2020. Association for Computational Linguistics.
- [6] Ratish Puduppully and Mark Steedman. Multi-document summarization with centroid-based pretraining, 2022.
- [7] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks, 2017.
- [8] Wen Xiao, Iz Beltagy, Giuseppe Carenini, and Arman Cohan. Primera: Pyramid-based masked sentence pre-training for multi-document summarization, 2022.

[9] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. Pegasus: Pretraining with extracted gap-sentences for abstractive summarization, 2020.

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
43	(1):	Despite its feasibility, self-	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	We propose a self-supervised learning framework for visual odometry (VO)	supervised VO still	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	that incorporates correlation of consecutive frames and takes advantage of	underperforms supervised ones.	0.0188 (prec)	0.0351 (prec)	0.0333 (prec)	0.0 (prec)	0.028 (prec)	0.0143 (prec)
	adversarial learning. Previous methods tackle self-supervised VO as a local	Apart from the effectiveness of	0.0361 (recall)	0.0482 (recall)	0.0602 (recall)	0.0 (recall)	0.0482 (recall)	0.012 (recall)
	structure from motion (SfM) problem that recovers depth from single image	direct supervision, a key reason is	0.0247 (f-1)	0.0406 (f-1)	0.0429 (f-1)	0.0 (f-1)	0.0354 (f-1)	0.0131 (f-1)
	and relative poses from image pairs by minimizing photometric loss between	that they focus mainly on	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	warped and captured images. As single-view depth estimation is an ill-posed	geometric properties @cite but	0.1242 (prec)	0.1217 (prec)	0.0993 (prec)	0.1591 (prec)	0.0972 (prec)	0.1127 (prec)
	problem, and photometric loss is incapable of discriminating distortion	pay little attention to the	0.2381 (recall)	0.1667 (recall)	0.1786 (recall)	0.0833 (recall)	0.1667 (recall)	0.0952 (recall)
	artifacts of warped images, the estimated depth is vague and pose is	sequential nature of the problem.	0.1633 (f-1)	0.1407 (f-1)	0.1277 (f-1)	0.1094 (f-1)	0.1228 (f-1)	0.1032 (f-1)
	inaccurate. In contrast to previous methods, our framework learns a compact	In these methods, only a few	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	representation of frame-to-frame correlation, which is updated by	frames (no more than 5) are	We propose a self-supervised	we present an unsupervised	"We present an unsupervised	In @cite, the authors proposed	@cite proposed an unsupervised	In @cite, a fully unsupervised
	incorporating sequential information. The updated representation is used for	processed in the network, while	learning framework for visual	learning framework for the task	learning framework for the task	an end-to-end learning	learning framework for the task	learning framework for
	depth estimation. Besides, we tackle VO as a self-supervised image generation	previous estimations are	odometry (VO) that incorporates	of monocular depth and camera	of monocular depth and camera	framework for monocular depth	of monocular depth and camera	monocular depth and camera
	task and take advantage of Generative Adversarial Networks (GAN). The	discarded and the current	correlation of consecutive	motion estimation from	motion estimation from	and camera motion estimation	motion estimation from	motion estimation from
	generator learns to estimate depth and pose to generate a warped target	estimation is made from scratch.	frames and takes advantage of	unstructured video sequences.	unstructured video sequences.	from unstructured video	unstructured video sequences.	unstructured video sequences
	image. The discriminator evaluates the quality of generated image with high-	Instead, the performance can be	adversarial learning. Previous	our method uses single-view	In common with recent work	sequences. In contrast to	Their method uses an end-to-	proposed. Unlike @cite and
	level structural perception that overcomes the problem of pixel-wise loss in	enhanced by taking geometric	methods tackle self-supervision	depth and multiview pose	[10, 14, 16], we use an end-to-	supervised methods, our	end learning approach with view	ocite which require only
	previous methods. Experiments on KITTI and Cityscapes datasets show that	relations of sequential	VO as a local structure from	networks, with a loss based on	end learning approach with view	framework learns to estimate	synthesis as the supervisory	monocular video sequences fo
	our method obtains more accurate depth with details preserved and predicted	observations into account.	motion (SfM) problem that	warping nearby views to the	synthesis as the supervisory	depth and pose by minimizing	signal. In contrast to the	training, our method is
	pose outperforms state-of-the-art self-supervised methods significantly.		recovers depth from single	target using the computed	signal. In contrast to the	photometric loss between	previous methods, our method	completely unsupervised and
	(2):		image and relative poses from	depth and pose.	previous work, our method is	warped and captured images.	learns a compact representation	can be applied independently
	We present an unsupervised learning framework for the task of monocular		image pairs by minimizing	the generator learns to estimate	completely unsupervised,	Comment:	of frame-to-frame correlation.	test time. In addition, we do no
	depth and camera motion estimation from unstructured video sequences. In		photometric loss between	depth and pose to generate a	requiring only monocular video	<ul> <li>Contrasts 2<sup>nd</sup> article's approach</li> </ul>		rely on a single view as the
	common with recent work [10, 14, 16], we use an end-to-end learning		warped and captured images. As	warped target image.	sequences for training. Our	(green) with 1st article's		supervisory signal, and we are
	approach with view synthesis as the supervisory signal. In contrast to the		single-view depth estimation is	the discriminator evaluates the	method uses single-view depth	approach (yellow)	information, Besides, we tackle	able to train the camera in a
	previous work, our method is completely unsupervised, requiring only		an ill-posed problem, and	quality of generated image with	and multiview pose networks,	- Clear sign of MDS	VO as a self-supervised image	supervised manner.
	monocular video sequences for training. Our method uses single-view depth		photometric loss is incapable of		with a loss based on warping	Words in red appeared in the		Comment:
	and multiview pose networks, with a loss based on warping nearby views to		discriminating distortion		nearby views to the target using	articles but cannot factuality		<ul> <li>Appears to contrast 2<sup>nd</sup> article</li> </ul>
	the target using the computed depth and pose. The networks are thus coupled		artifacts of warped images, the	pixel-wise loss in previous	the computed depth and pose.	cannot be verified based on		(green) with the main one
	by the loss during training, but can be applied independently at test time.		estimated depth is vague and		The networks are thus coupled	inputs alone	generator learns to estimate	- But despite the paragra
	Empirical evaluation on the KITTI dataset demonstrates the effectiveness of		pose is inaccurate. In contrast to	experiments on the KITTI and	by the loss during training, but		depth and pose to generate a	structure, the red part actua
	our approach: 1) monocular depth performs comparably with supervised		previous methods, our	Cityscapes datasets show that	can be applied independently at		warped target image. The	came from 2nd article
	methods that use either ground-truth pose or depth for training, and 2) pose		framework learns a compact	our method obtains more	test time. Empirical evaluation		discriminator evaluates the	
	estimation performs favorably compared to established SLAM systems under		representation of frame-to-	accurate depth with details	on the KITTI dataset		quality of generated image with	
	comparable input settings.		frame correlation, which is	preserved and predicted pose	demonstrates the effectiveness		high-level structural perception	
	· · · · -		updated by incorporating	outperforms state-of-the-art	of our approach: 1) monocular		that overcomes the problem of	
			sequential information. The	self-supervised methods	depth performs comparably with		pixel-wise loss in previous	
			updated representation is used		supervised methods that use		methods. Experiments on KITTI	
			for depth estimation. Besides,	Comment:	either ground-truth pose or		and Cityscapes datasets show	
			we tackle VO as a self-		depth for training, and 2) pose		that our method obtains more	
			Supervised image generation	- First section came from 2 <sup>nd</sup>	estimation performs favorably		accurate depth with details	
				article, while parts in red came				
			task and take advantage of	from the 1st article	compared to established SLAM		preserved and predicted pose	
			Generative Adversarial Networks	- No clear sign of MDS as the	systems under comparable input		outperforms state-of-the-art	
			(GAN). The generator learns to	text does not distinguish the	settings."		methods significantly.	
			estimate depth and pose to	sources	Comment:		Comment:	
			generate a warped target image.	Sources	- Copied from the second article		<ul> <li>Contrasts 2<sup>nd</sup> article's approach</li> </ul>	
		1	The discriminator evaluates the	I	only		(green) with 1st article's	I
		1	quality of generated image				approach (yellow)	1
		1	Comment:	I	- No indication of MDS		- Clear sign of MDS	I
		1	- Copied the first tokens only				<ul> <li>No wrong or unverifiable facts,</li> </ul>	
		1	- No indication of MDS				but copying quite a lot	[
ı		1	The state of the s				and any management of the	1
1		I						1
		1	1	I .	1	l .	1	1

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
483	(1):	In a slightly different problem,	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
1	This paper builds upon the current methods to increase their capability and	where a NN is used to reconstruct	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
1	automation for 3D surface construction from noisy and potentially sparse	the shape of a 3D object from its	0.0643 (prec)	0.0743 (prec)	0.0505 (prec)	0.082 (prec)	0.0806 (prec)	0.0843 (prec)
1	point clouds. It presents an analysis of an artificial neural network surface	shading in a 2D @cite . show	0.0932 (recall)	0.0932 (recall)	0.0932 (recall)	0.0424 (recall)	0.0424 (recall)	0.0593 (recall)
1	regression and mapping method, describing caveats, improvements and instification for the different approach.	from experiment that	0.0761 (f-1)	0.0827 (f-1)	0.0655 (f-1)	0.0559 (f-1)	0.0556 (f-1)	0.0697 (f-1)
		quantitative improvement does	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	(2): Shape-from-shading (SFS) methods tend to rely on models with few	not necessarily lead to quantitative improvement. This is	0.1337 (prec) 0.1933 (recall)	0.1409 (prec) 0.1765 (recall)	0.1187 (prec) 0.2185 (recall)	0.2581 (prec) 0.1345 (recall)	0.3016 (prec) 0.1597 (recall)	0.2143 (prec) 0.1513 (recall)
	parameters because these parameters need to be hand-tuned. This limits the	something to consider when	0.1933 (recall) 0.1581 (f-1)	0.1765 (recall) 0.1567 (f-1)	0.2185 (recall) 0.1538 (f-1)	0.1345 (recall) 0.1768 (f-1)	0.1597 (recall) 0.2088 (f-1)	0.1513 (recall) 0.1773 (f-1)
	number of different cues that the SFS problem can exploit. In this paper, we	using a 'black box' function like a	0.1581 (I-1) Summary:	Summary:	0.1538 (I-1) Summary:	Summary:	Summary:	Summary:
	show how machine learning can be applied to an SFS model with a large	neural network, especially where	This paper builds upon the	this paper builds upon the	In this article, we're going to	In @cite, a hemisphere is	In @cite, the authors propose a	In @cite, an SFS model is used to
	number of parameters. Our system learns a set of weighting parameters that	there could be some information	current methods to increase	current methods to increase	explain how machine learning	rendered using a known	method to learn a set of	learn the intensity of each pixel
	use the intensity of each pixel in the image to gauge the importance of that	loss. In this regard we must	their capability and automation	their capability and automation	can be applied to an artificial	illumination direction. The	parameters that use the	in the image. The intensity of the
	pixel in the shape reconstruction process. We show empirically that this leads	ensure that the final model is	for 3D surface construction from	for 3d surface construction from	neural network surface	isophotes in the resulting	intensity of each pixel in the	image is used as the input to the
	to a significant increase in the accuracy of the recovered surfaces. Our learning	representative of the ground	noisy and potentially sparse	noisy and potentially sparse	regression and mapping	reflectance map are then	image to gauge the importance	SFS model. The weights can also
	approach is novel in that the parameters are optimized with respect to actual	truth and not only rely on an	point clouds. It presents an	point clouds.	method, describing caveats,	modelled using Gaussian	of that pixel in the shape	be optimized for a particular
	surface output by the system. In the first, offline phase, a hemisphere is	error measure. It is suggested	analysis of an artificial neural	it presents an analysis of an	improvements and justification	mixtures to obtain a parametric	reconstruction process. They	input image. In this work, we
	rendered using a known illumination direction. The isophotes in the resulting	that more research must be done	network surface regression and	artificial neural network surface	for the different approach. This	representation of the isophotes.	show empirically that this leads	show how machine learning can
	reflectance map are then modelled using Gaussian mixtures to obtain a	for 3D surface quality metrics	mapping method, describing	regression and mapping	paper builds upon the current	This is then used in the second	to a significant increase in the	be applied to a SFS model with a
	parametric representation of the isophotes. This Gaussian parameterization is	@cite . Visual quality will be	caveats, improvements and	method, describing caveats,	methods to increase their	phase to learn intensity-based	accuracy of the recovered	large number of parameters. We
1	then used in the second phase to learn intensity-based weights using a	assessed in the method	justification for the different	improvements and justification	capability and automation for 3D	weights using a database of 3D	surface. The weights can also be	also show how the weighting
	database of 3D shapes. The weights can also be optimized for a particular	presented here alongside	approach.     Shape-from-	for the different approach.	surface construction from noisy	shapes. The weights can also be	optimized for a particular input image.	parameters can be used to learn
	input image.	quantitative results in the absence of quality metrics.	shading (SFS) methods tend to rely on models with few	methods tend to rely on models	and potentially sparse point clouds. It presents an analysis of	optimized for a particular input image.	Comment:	intensity-based weights using a database of 3D shapes.
		absence or quality metrics.	parameters because these	with few parameters because	an artificial Neural Network	Comment:	- Summarized the 2nd article	Comment:
			parameters need to be hand-	these parameters need to be	surface regression and mapped	- Summarized the 2 <sup>nd</sup> article	only	- Summarized the 2 <sup>nd</sup> article
			tuned. This limits the number of	hand-tuned.	method, describing caveat,	only	The first article is probably too	only
			different cues that the SFS	this limits the number of	improvements, and justification	- The first article is probably too	short to provide sufficient	- The first article is probably too
			problem can exploit. In this	different cues that the sf	for a different approach. In this	short to provide sufficient	inputs?	short to provide sufficient
			paper, we show how machine	problem can exploit. in this	paper, we show how machine	inputs?	- Weak indication of MDS	inputs?
			learning can be applied to an SFS	paper	learning is applied to an SFS	- Weak indication of MDS		- Weak indication of MDS
			model with a large number of	, we show how machine learning	model with a large number of			
			parameters. Our system learns a	can be applied to an sf model	parameters. Our system learns a			
			set of weighting parameters that	with a large number of	set of weighting parameters that			
			use the intensity of each pixel in	parameters.	use the intensity of each pixel in			
			the image to gauge the	our system learns a set of	the image to gauge the			
			importance of that pixel in the	weighting parameters that use	importance of that pixel in the			
			shape reconstruction process.  We show empirically that this	the intensity of each pixel in the image to gauge the importance	shape reconstruction process. We show that this leads to a			
			leads to a significant increase in	of that pixel in the shape	significant increase in the			
			the accuracy of the recovered	reconstruction process.	accuracy of the recovered			
			surfaces. Our learning approach	we show empirically that this	surfaces. Our learning approach			
			is novel in that the parameters	leads to a significant increase in	is novel in that the parameters			
			are optimized with respect to	the accuracy of the recovered	are optimized with respect to			
			actual surface output by the	surfaces.	actual surface output by the			
			system. In the first	Comment:	system. In the first, offline			
			Comment:	- Copied the first tokens only,	phase, a hemisphere is rendered			
			- Copied the first tokens only	and even with errors (red)	using a known illumination			
			- No indication of MDS	- No indication of MDS	direction. The isophotes in the			
			1		resulting reflectance map are then modelled using Gaussian			
			1		mixtures to obtain a parametric			
					representation of the isophotes.			
					This Gaussian parameterization			
			1		is then used in the second phase			
			1		to learn intensity-based weights			
			1		using a database			
			1		Comment:			
			1		- Initial parts shows some			
			1		contrasts between 1st (green)			
			1		and 2nd yellow articles			
			1		- Some sign of MDS, but goes on			
			1		to just copy everything else from 2nd article (red)			
1			1		Irom 2 <sup>rd</sup> article (red)			
1			1					
			L					

556 (1):  1a disaster situation, first responders need to quickly acquire situational awareness and prioritize response based on the need, resources available and properties of the properties	Rouge scores: Rouge scores: - Rouge 2: - Rouge 2:	Rouge scores:	
	- Pougo 3:		Rouge scores:
awareness and prioritize response based on the need, resources available and replace newswire for breaking 0.0419 (prec) 0.049 (prec)		- Rouge 2:	- Rouge 2:
	0.0625 (prec) 0.0431 (prec)	0.0536 (prec)	0.0111 (prec)
impact. Can they do this based on digital media such as Twitter alone, or news @cite . They studied a 0.0583 (recall) 0.0417 (recall)	0.0833 (recall) 0.0417 (recall)	0.05 (recall)	0.0083 (recall)
newswire alone, or some combination of the two? We examine this question period of 77 days in 2011 during 0.0488 (f-1) 0.045 (f-1)	0.0714 (f-1) 0.0424 (f-1)	0.0517 (f-1)	0.0095 (f-1)
in the context of the 2015 Nepal Earthquakes. Because newswire articles are which 27 events occurred. The - Rouge L: - Rouge L:	- Rouge L: - Rouge L:	- Rouge L:	- Rouge L:
longer, effective summaries can be helpful in saving time yet giving key biggest disasters in this event-set 0.1548 (prec) 0.1845 (prec)	0.1366 (prec) 0.1709 (prec)	0.177 (prec)	0.1868 (prec)
content. We evaluate the effectiveness of several unsupervised summarization are: an airplane crash with 43 0.2149 (recall) 0.157 (recall)	0.1818 (recall) 0.1653 (recall)	0.1653 (recall)	0.1405 (recall)
techniques in capturing key content. We propose a method to link tweets deaths, and a magnitude 5.8 0.1799 (f-1) 0.1696 (f-1)	0.156 (f-1) 0.1681 (f-1)	0.1709 (f-1)	0.1604 (f-1)
written by the public and newswire articles, so that we can compare their key earthquake in Virginia that Summary: Summary:	Summary: Summary:	Summary:	Summary:
characteristics: timeliness, whether tweets appear earlier than their caused infrastructural damage. In a disaster situation, first in a disaster situation			In @cite, the authors found that
corresponding news articles, and content. A novel idea is to view relevant  None of these disasters, bad as responders need to quickly responders need to a quick			newswire providers report the
tweets as a summary of the matching news article and evaluate these they are, rise to the level of the acquire situational awareness acquire situational awareness			same events as newswire
summaries. Whenever possible, we present both quantitative and qualitative  Nepal Earthquake(s) of 2015 in and prioritize response based on a manifest of the prioritize response based on the prioriti			providers, in addition to a long
evaluations. One of our main findings is that tweets and newswire articles which almost 10,000 lives were the need, resources available the need, resources available			tail of minor events ignored by
provide complementary perspectives that form a holistic view of the disaster lost. They collected a large and impact. Can they do this and impact.	extent to which news reporting traditional newsw		mainstream media. However,
situation. dataset of tweets and news based on digital media such as can they do this			they found that Twitter does not
(2): articles, but then eliminated a Twitter alone, or newswire media such as Tv			report the same event as
Twitter is often considered to be a useful source of real-time news, potentially large collection of tweets based alone, or some combination of newswire alone,			newswire, which suggests that
replacing newswire for this purpose. But is this true? In this paper, we examine on clustering. More elimination of the two? We examine this combination of the two? We examine this			the importance of news
the extent to which news reporting in newswire and Twitter overlap and tweets led to only 97 linked question in the context of the we examine this			reporting in a newswire setting
whether Twitter often reports news faster than traditional newswire providers. tweet-news article pairs, which is 2015 Nepal Earthquakes. context of the 20			comes from increased event
In particular, we analyse 77 days worth of tweet and newswire articles with a small dataset. Because newswire articles are Earthquakes.	articles with respect to both events. Their resu	ts indicate events. Their results indicate	coverage, not timeliness of
respect to both manually identified major news events and larger volumes of longer, effective summaries can we propose a m	had to link manually identified major news that Twitter repor	ts the same that Twitter reports the same	reporting. The authors conclude
automatically identified news events. Our results indicate that Twitter reports be helpful in saving time yet weets written be	he public and events and larger volumes of events as newswir	e providers, in events as newswire providers, in	that Twitter is a useful source of
the same events as newswire providers, in addition to a long tail of minor giving key content. We evaluate newswire article	so that we automatically identified news addition to a long	tail of minor addition to a long tail of minor	real-time news, but it is not clear
events ignored by mainstream media. However, contrary to popular belief, the effectiveness of several can compare the	key events. Our results indicate that events ignored by	mainstream events ignored by mainstream	whether this is the case for both
neither stream leads the other when dealing with major news events, unsupervised summarization characteristics:	eliness, Twitter reports the same events media. However, o	ontrary to media. However, contrary to	newswire and newswire
indicating that the value that Twitter can bring in a news setting comes techniques in capturing key	pear earlier as newswire providers, in popular belief, nei	ther stream popular belief, neither stream	providers.
predominantly from increased event coverage, not timeliness of reporting. content.We propose a method than their corres			Comment:
to link tweets written by the articles, and con			- Weak indication of MDS, with
public and newswire articles, so evaluate the effe			the 2nd article presented as not
that we can compare their key several unsuper			the main article
characteristics: timeliness, summarization t			- The first red part is factually
whether tweets appear earlier capturing key co			wrong, while the second part is
than their corresponding news Comment:	indicating that the value that not timeliness of r		not implied from the inputs
articles, and content. A novel		Comment:	not implied from the inputs
idea is to view relevant tweets only but with			
as a summary of the matching two sentences			
news article and evaluate these - No indication of			
summaries. Whenever possible,	of our main findings is that one, i.e. weak	indication of - Weak indication of MDS	
we present both quantitative	tweets and newswire provide MDS		
and qualitative evaluations. One	complementary perspectives		
of our main findings is that	that form a holistic view of the		
tweets and newswire are	disaster situation.		
complementary perspectives	Comment:		
that form a holistic view of	- Except for the last sentence		
	(yellow), all parts are from 2nd		
	article with some paraphrasing		
	- Summary presented the		
	studies as the same one, i.e.		
	no indication of MDS		

that the length of the entities in significant consequences for our understanding of natural and man-made various datasets follows a family of scale-free power law distributions. The part of the distribution representing large but rare events—and by the neithy here broadly includes the neithy her	No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
Secure blooks than of the temporary to the care in reading and extremest religion.  In the care of the	599								
The company from the process growing of the company from									
personal part of the management of the managemen									
Anguer in commonstrate growth people devices between the property and the confidence of the confidence									
be number of worth it an artific the control minimal by the concent mental in public that the control mental in the control mental in public that the control mental in the control									
posses the care free groups for all the set of feedings and program for the control of the contr									
owners we special the phenomenous of pager levels in early large (1) and the properties of according to the properties of ac									
The page properties a discovery but the length of the entire that the length of									
that the longs and part of the written is a complicated by the light distribution the court of the part of the distribution of the control of the distribution of the court of th		principle of least effort in communication and the preferential mechanism.		Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
works distinct follows family and individuals in comparing sings but are event—ead by the implication representing sings but are event—ead by the implication of the distinction representing sings but are event—ead by the implication of the distinction representing sings but are event—ead by the implication of the distinction representing sings but are event—ead by the implication of the distinction representing sings but are event—ead by the implication of the distinction representing sings but are event—ead by the implication of the distinction of the second single sin		(2):		This paper presents a discovery	this paper presents a discovery	Power-law distributions occur in	In @cite, the authors present a		In @cite, the authors present a
In completion the court in the set of the desirable of classification of the court per the set of desirable of the set of the court per the set of desirable of the court per the set of desirable of the court per the set of desirable of the court per the set of the court per the set of desirable of the court per the set of the s									statistical framework for
accomplicated by the large fluctuations that occurs in the sale of the distributions of the d									discerning and quantifying
— the part of the districtions representing large but any control and early early proposed to design find process on which power law individual for the control of the cont									power-law distributions. The
officulty of destribing the practice with proposed with objective holds.  Commonly and method fire analysis provided and the experience approaches and the proposed and the prop									authors present a method that
Commonly used methods for analysing power-law data, such as least-squares finding, campained authorized increases and interest of the properties of the companies and contained and contained of whether the data down a power for all these displayable processing and existent states of the data down and									combines maximum-likelihood fitting with the Kolmogorov-
fitting_care produces substantially inscruzine extensions of parameters for power law distributions, and one mass where some methods return or any even in the production of the parameters of all returns and productions of the parameters of a section of the production of the parameters of a section of the parameters of the parameter									Smirnov (KS) statistic and
sower-law distributions, and even to now whether such methods return of courts were such methods of the proposed method to the court of the approach with production the court of the approach with a production of the approach with production the court of the approach with a production the court of the approach with a production the court of the approach with production the production of the approach with a production the court of the production of the approach with a production the court of the production of the									likelihood ratios. The authors
security answer they are still unardisticatory section by each of the data days appear the set of the section o									demonstrate that their method
the number of words in an early. The procedure of the group and appropriate process are supported to the process of the agronal part of the distribution of grower laws of the distribution of grower laws of the state of the consistent with the distribution of grower laws of the state of the					we explain the phenomenon of				can be used with synthetic data
amount of the communication and the granter communication and the granter communication and the preferential mechanism. Unfortunately, the detection or down from your desirable of the distributions or down from the data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data, while in others the power law is ruled out.  The data and the standard of the data, while in others the power law is ruled out.  The data and the standard of the data, while in others the power law is ruled out.  The data and the standard of the data, while in others the power law is ruled out.  The data and the standard of the data and the data and the data and the data formed and the data		of whether the data obey a power law at all. Here we present a principled		areas. The entity length denotes	power laws in entity length by	the distribution representing	synthetic data and give critical	synthetic data and give critical	and give critical comparisons to
with goodness of efficiency of the efficiency of									previous approaches. However,
possess the scale free property and how we will define disease of the time of the second control of the second									the authors do not provide a
and have well-defined many and proposed methods to be grown beautiful to the processor and the proposed methods to the consistent with the data, while in others the power law is ruled out.  and have well-defined many and the processor and the preferential mechanism. [IIII] Power law distributions occur in measure and have againful and many and preferential mechanism. [IIII] Power law distributions occur in measure and have againful and many and preferential of the data, while in others the power law is ruled out.  and have well-defined many and the preferential mechanism. [IIII] Power law distributions occur in measure and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of natural and many measures and have againfluint consequences for our understanding of the definition of the									quantitative comparison of their
and finite variances. We explain the phenomenon of power law is richt before discipline, each of which has been competented to follow a power- law distribution in some cases we find these conjectures to be consistent with the data, write in other the power law is rised out.  We distribution of scientific interest and have significant to an amount of phenomenon an among phenomenon.  Unfortunately, the detection and characterization of power laws to all the standard and the stand									method with the KKS method,
the phenomenon of power laws in entity length by the principal of laster february is the principal of laster february is the province of laster february is the dark, while in others the power law is ruled out.  If the dark, while in others the power law is ruled out.  If the dark, while in others the power law is ruled out.  If the dark is the power law is ruled out.  If the phenomenon of power laws is ruled out.  If the power law is ruled ou									empirical comparison of their
in entity length by the principle of lear effect in community and the preference and the preference and the preference and the preference and the segment of the preference and the preference and the preference and the segment of the preference and the preference and the preference and the segment of the preference and									results with KKS methods.
of least effort in communication and the preferential mechanism. [III] Proper-law distributions corum in many the professor and how to ignificant correct and how to ignificant correct and man-made phenomena. Unfortunately, the detection and complicated by the Ingerial Complex Index									
and the preferential mechanism. [III] Prover-law distributions occur in many students of classific the consequences for our understanding of natural and man-made phenomena. Unfortunately, the destroy on several and man-made phenomena. Unfortunately, the destroy of the distribution represented large but relevant on the distribution representing large but rere events—and by the distribution representing large but rere events—and by the distribution representing large but rere events—and by the distribution of power-law of the arrange over which power-law Comment:  - Copied the first tokens only—8. No indication of MDS  - No lend grain of the distribution of the control of the distribution represented as a first development of the distribution representing large but are events—and by the distribution represents are large over which power-law comments.  - Copied the first tokens only—8. No local argue of MDS as the constant of the distribution represents and the preferential mechanism.  - Copied the first tokens only—8. No local argue of MDS as the constant large development of the distribution represents and the preference of the distribution of the distribution of the distribution of the distribution represents and the distribution of the distribution represents and the distribution of the distribution of the distribution represents and the distribution of the distribution of the distribution of the distribution of the		the data, while in others the power law is ruled out.		of least effort in communication	likelihood ratios.	distributions. The concept of		they find these conjectures to be	- All parts are from 2nd article.
distributions occur in many situations of social methods are grown as entered and have significant understanding of natural and man-made phenomena. Unfortunately, the detection and characterization of power laws is complicated by the detection and characterization of power laws is complicated by the detection and characterization of power laws is complicated by the destroyation of the distribution representing large but rare events—and by the distribution representing large but rare events—and by the distribution of MDS  Copied the first between only:  No indication of MDS  which is between only:  No indication o				and the preferential	we evaluate the effectiveness of	entity here broadly includes the		consistent with the data, while	Similar to the other tuned
stuations of scientific interest and above significant or consequences for our understanding of natural and man-made phenomena. Understanding for natural and man-made phenomena. Understanding for natural and complicated by the large fluctuations that one science was a complicated by the large fluctuations that one presenting large but rare events—and by the distribution—the part of the distribution—the part of the distribution of presenting large but rare events—and by the distribution of presenting large but rare events—and by the distribution of presenting large but rare events—and by the distribution of the presentation of the button of the presentation of the button of the bu					the approach with tests on				
and have significant consequences for our understanding of natural and man-made phenomena. Unfortunately, the detection and characterization is been conjugated by the ingre- of the distribution representing large but are seen two strated comments:  - Comment:  - Opper the first bleven only is No enducation of MIDS.  - No clear sign of MIDS as the source law should be sourced as a strategisted on matural language processing and related and processing and related anguage processing anguage process									- Red part contains hallucinated
consequences for our understanding of natural and man-made phenomena.  Unfortunately, the detection and characterization of power the detection and characterization of power the power law is complicated by the large fluctuations that occur in the tail of the distributions representing a large but rare events—and by the difficulty of the power-law behavior comment:  - Copied the first tokens only - No indication of MDS  No indication of MDS  - No indication of MDS  - No indication of MDS  - Region of MDS  - Regi									information
understanding of natural and man-mande phenomena. Unfortunately, the detection and characterisation of power laws is complicated by the large fluctuations of power laws is complicated by the large fluctuation representing large but rare events—and by the difficulty interest to be consistent with the data, while it the data, while it is described by the difficulty interest to be consistent with behavior.  Comment:  Comment:								- All parts are from 2 <sup>nd</sup> article.	
man-made phenomena.  Unfortunately, the detection and characterization of power laws is complicated by the large fluctuations that occur in the tail of the distribution representing large but rare events—and by the large fluctuations that occur in the tail of the distribution representing large but rare events—and by the large fluctuations that occur in the tail of the distribution representing the large but rare events—and by the large fluctuation of the distribution representing the large but rare events—and by the large over which power-law behavior comment:  - Copie dut finite to takens only - No indication of MIDS  - No indication of									
Unfortunately, the detection and characteristication of power laws is complicated by the large fluctuations that cour in the tail of the distribution- representing large but rarse event—and by the difficulty of dentifying the range over which power-law behavior:  Comment:  - Copine the first tokens only - No institution of MIDS  - No institution of MIDS  - No institution of MIDS  - First section came from 1 and the perferential mechanism.  - First section came from 1 and the preferential mechanism.  - First section came from 1 and the preferential mechanism.  - From the 2-4 article - No clear sing of MIDS as the text does not distinguish the sources.  - Oping the first tokens only - No clear sing of MIDS as the text does not distinguish the sources.  - Oping the first tokens only - No clear sing of MIDS as the text does not distinguish the sources.  - Oping the first tokens only - No clear sing of MIDS as the text does not distinguish the sources.  - Oping the first tokens only - No clear sing of MIDS as the text does not distinguish the sources.  - No clear sing of MIDS as the text does not distinguish the sources.  - No clear sing of MIDS as the text does not be lightly as the sources.  - No clear sing of MIDS as the text does not be lightly as the source of the sourc									
Complicated by the large fluctuations that occur in the tail of the distribution—the part of the distribution—the part of the distribution representing the green that are events—and by the difficulty of identifying the principle the range over which power-law behavior  Comment:  - Conjenent: - No indication of MDS  -									
in commentation of MDS  In com				characterization of power laws is	which has been conjectured to	distributions in entity length		one, i.e. weak indication of	
of the distribution—the part of the distribution representing large but rare events—and by the difficulty of identifying the range over which power-law behavior  Comment:  - Copied the first tokens only - No indication of MDS  No indication of MDS  - No indication of MDS  - No indication of MDS  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sou					follow a power-law distribution.			MDS	
the distribution representing large but rare event—and by the difficulty of identifying the range over which power-law behavior  Comment:  - Copied the first tokens only  - No locar sign of MDS as the text does not distinguish the sources  - No locar sign of MDS as the text does not distinguish the sources  - No indication of MDS  - No indi									
large but rare events—and by the difficulty of identifying the range over which power-law behavior  Comment:  - Copied the first tokens only - No indication of MDS  - No indi									
comment: - Copied the first tokens only - No indication of MOS  No indication of MOS  - No indication									
First section came from 1s article, while parts in red came - Copied the first tokens only - Opied the first tokens only - No indication of MDS  - No indication of MDS  - No indication of MDS  - No clear sign of MDS as the does not distinguish the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of MDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sources  - No clear sign of NDS as the sour									
Comment:  - Copied the first tokens only - No indication of MDS  No indication of MDS  - No indication									
from the 2-w article - Copied the first tokens only - No indication of MDS  No clear sign of MDS as the text does not distinguish the sources  and an						Power-law distributions are not			
- Copied the first tokens only  No indication of MDS as the text does not distinguish the sources  - No clear sign of MDS as the text does not distinguish the sources  - No indication of MDS as the text does not distinguish the sources  - Indicating the standard fistory, such as the source of th						limited to natural phenomena.			
text does not distinguish the sources  text does not distinguish the sources  text does not distinguish the sources are a made climate change, and natural bitstry, such as text and a made climate change and natural part of the source, they can be difficult to detect due to the large Comment:  - Only the 1" article (yellow) used in describing the study, while 2" article (green) only used to provide general context  - Red part is also from 2" article and somehow appeared again despite the idea is also covered in the green part									
sources  required https://www.min.  returnal pas and other next.trial phenomenan. Nouver, they call be difficult to letter due to this  Comment:  - Only the 1st article (yellow) used in describing the study, while 2st article (gene) only used in 2st article (gene) only used to provide general context  - Red part is also from 2st article and somehow appeared again despite the idea is also covered in the green part				- No indication of MDS					
the promote a flow of the control o									
incomments towards they do to the comments of									
in difficult to defect due to the transport of tra									
Comment:  - Only the 1st article (yellow) used in describing the study, while 2nd article (general context as the study of									
- Only the 1st article (yellow) used in describing the study, while 2nd article (green) only used to provide general context - Red part is also from 2st article and somehow appeared again despite the idea is also covered in the green part						large			
used in describing the study, while 2 <sup>nd</sup> article (green) only used to provide general context - Red parts also from 2 <sup>nd</sup> article and somehow appeared again despite the idea is also covered in the green part						Comment:			
used in describing the study, while 2 <sup>nd</sup> article (green) only used to provide general context - Red parts also from 2 <sup>nd</sup> article and somehow appeared again despite the idea is also covered in the green part						- Only the 1s article (vellow)			
while 2 <sup>nd</sup> article (green) only used to provide general context - Ned part is also from 2 <sup>nd</sup> article and somehow appeared again despite the idea is also covered in the green part									
used to provide general context - Red part is also from 2 <sup>nd</sup> article and somehow appeared again despite the idea is also covered in the green part									
context  - Red part is also from 2 <sup>rd</sup> article and somehow appeared again despite the idea is also covered in the green part									
and somehow appeared again despite the idea is also covered in the green part									
despite the idea is also covered in the green part									
covered in the green part									
- no deal sign of mus									
						- No clear sign or wide			

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
620	(1):	The improved methods that we	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	In this work we investigate the reasons why Batch Normalization (BN)	propose are also closely related	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	improves the generalization performance of deep networks. We argue that	to variational drop -out @cite as	0.0068 (prec)	0.0351 (prec)	0.022 (prec)	0.0156 (prec)	0.0 (prec)	0.0 (prec)
	one major reason, distinguishing it from data-independent normalization	discussed below. We give a new	0.0303 (recall)	0.0606 (recall)	0.1212 (recall)	0.0303 (recall)	0.0 (recall)	0.0 (recall)
	methods, is randomness of batch statistics. This randomness appears in the	interpretation to variational	0.0111 (f-1)	0.0444 (f-1)	0.0372 (f-1)	0.0206 (f-1)	0.0 (f-1)	0.0 (fecall)
	parameters rather than in activations and admits an interpretation as a	dropout and apply it in	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	practical Bayesian learning. We apply this idea to other (deterministic)	combination with normalization	0.0676 (prec)	0.1207 (prec)	0.0601 (prec)	0.0923 (prec)	0.1042 (prec)	0.0795 (prec)
	normalization techniques that are oblivious to the batch size. We show that	techniques.	0.2941 (recall)	0.2059 (recall)	0.3235 (recall)	0.1765 (recall)	0.1471 (recall)	0.2059 (recall)
	their generalization performance can be improved significantly by Bayesian		0.1099 (f-1)	0.1522 (f-1)	0.1014 (f-1)	0.1212 (f-1)	0.122 (f-1)	0.1148 (f-1)
	learning of the same form. We obtain test performance comparable to BN		Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	and, at the same time, better validation losses suitable for subsequent output		In this work we investigate the	we investigate a local	In this article, we investigate the	In @cite, the authors propose a	In <mark>@cite</mark> , a <mark>local</mark>	In @cite, a local
	uncertainty estimation through approximate Bayesian posterior.		reasons why Batch	reparameterizaton technique for	reasons why Batch	local reparameterization that	reparameterization is used to	reparameterization is used to
	(2):		Normalization (BN) improves the	greatly reducing the variance of	Normalization (BN) improves the	translates uncertainty about	reduce the variance of stochastic	reduce the variance of stochastic
	We investigate a local reparameterizaton technique for greatly reducing the		generalization performance of	stochastic gradients for	generalization performance of	global parameters into local	gradients for variational	gradients for variational
	variance of stochastic gradients for variational Bayesian inference (SGVB) of a		deep networks. We argue that	variational bayesian inference	deep networks. We argue that	noise that is independent across	Bayesian inference (SGVB) of a	Bayesian inference (SGVB) of a
	posterior over model parameters, while retaining parallelizability. This local		one major reason, distinguishing	( sGVB ) of a posterior over	one major reason, distinguishing	datapoints in the minibatch. The	posterior over model	posterior over model
	reparameterization translates uncertainty about global parameters into local		it from data-independent	model parameters, while	it from data-independent	authors show that this method	parameters, while retaining	parameters, while retaining
	noise that is independent across datapoints in the minibatch. Such		normalization methods, is	retaining parallelizability.	normalization methods, is	can be improved significantly by	parallelizability. This method	parallelizability. In contrast, our
	parameterizations can be trivially parallelized and have variance that is		randomness of batch statistics.	our method allows inference of	randomness of batch statistics.	Bayesian learning of the same	allows inference of more flexibly	method does not require
1	inversely proportional to the mini-batch size, generally leading to much faster	1	This randomness appears in the	more flexibly parameterized	This randomness appears in the	form. In @cite the authors	parameterized posteriors;	parallelizability, and we do not
1	convergence. Additionally, we explore a connection with dropout: Gaussian	1	parameters rather than in	posteriors ; specifically, we	parameters rather than in	propose variational dropout, a	specifically, the dropout rates	need parallelizability in our
1	dropout objectives correspond to SGVB with local reparameterization, a scale-	1	activations and admits an	propose variational dropout. a	activations and admits an	generalization of Gaussian	are learned, often leading to	method, which is a key
1	invariant prior and proportionally fixed posterior variance. Our method allows	1			interpretation as a practical	dropout where the dropout	better models.	advantage of SGVB over SGVB.
1		1	interpretation as a practical	generalization of gaussian				
	inference of more flexibly parameterized posteriors; specifically, we propose	1	Bayesian learning. We apply this	dropout where the dropout	Bayesian learning. We apply this	rates are learned, often leading	Comment:	In addition, our method can be
	variational dropout, a generalization of Gaussian dropout where the dropout		idea to other (deterministic)	rates are learned, often leading	idea to other (deterogeneous)	to better models.	<ul> <li>All parts are from 2<sup>nd</sup> article.</li> </ul>	seen as a generalization of SGVB
	rates are learned, often leading to better models. The method is demonstrated		normalization techniques that	to better models.	normalization techniques that	Comment:	- However, it is represented as a	where the dropout rates are
	through several experiments.		are oblivious to the batch size.	Comment:	are oblivious to the batch size.	<ul> <li>Summarizes the 1st (yellow),</li> </ul>	separate article from the main	learned, often leading to better
			We show that their	- Copied only from the 2nd	We show that their	and then the 2nd (green) article	one, i.e. weak indication of	models, and the variance is
			generalization performance can	article, specifically the first and	generalization performance can	- Shows signs of MDS, but	MDS	inversely proportional to the
			be improved significantly by		be improved significantly by	factuality is a concern since the		mini-batch size.
			Bayesian learning of the same	penultimate sentences.	Bayesian learning of the same	red part is actually from the		Comment:
			form. We obtain test	- No sign of MDS	form. We obtain test	2 <sup>nd</sup> article		- Contains mostly information
			performance comparable to BN	- Errors introduced quite	performance comparable to BN			from the 2 <sup>nd</sup> model (green)
			and, at the same time, better	randomly (red)	and, at the same time, better			- However, it pretends to be
			validation losses suitable for	randonny (red)	validation losses suitable for			doing an MDS (red) though the
					subsequent output uncertainty			information is not directly
			subsequent output uncertainty		estimation through approximate			
			estimation through approximate					implied and contains logical
			Bayesian posterior.    We		Bayesian posterior. Additionally,			inconsistencies ("key
			investigate a local		we explore a connection with			advantage of SGVB over
			reparameterizaton technique for		dropout: Gaussian dropout			SGVB")
			greatly reducing the variance of		objectives correspond to SGVB			
			stochastic gradients for		with local reparameterization, a			
			variational Bayesian inference		scale-invariant prior and			
			(SGVB) of a posterior over model		proportionally fixed posterior			
			parameters, while retaining		variance. Our method allows			
		1	parallelizability. Our method		inference of more flexibly			
			allows inference of a		parameterized posteriors;			
1		1	generalization		specifically, we propose			
		1	Comment:		variational dropout, a			
1		1	- Copied the first tokens only		generalization of Gaussian drop			
			- No indication of MDS		out where the dropout rates are			
1		1			learned, often leading to better			
1		1	1		models. The method is			
		1	1					
		1			demonstrated through several			
1		1	1		experiments. The Method is			
1		1	1		demonstrated through multiple			
		1	1		experiments.			
		1	1		Comment:			
		1	1		- Initial parts copied from the 1st			
		1	1		article (yellow), followed by			
		1	1					
		1	1		copying from the 2 <sup>nd</sup> article			
1		1	1		(green)			
		1			- No sign of MDS as the two are			
1		I	l .		presented as concerning the			
		1	1		same study			
		1	1		- A made-up word is randomly			
1		1	1		introduced for no apparent			
		1	1		reason (red)			
		1						
1	1	1	1					

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
3157	(1):	The success of word embedding	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	In this paper, we study a novel approach for named entity recognition (NER)	@cite encourages researchers to	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	and mention detection in natural language processing. Instead of treating NER	focus on machine-learned	0.0316 (prec)	0.0303 (prec)	0.035 (prec)	0.0685 (prec)	0.0 (prec)	0.049 (prec)
	as a sequence labelling problem, we propose a new local detection approach,	representation instead of heavy	0.0463 (recall)	0.037 (recall)	0.0648 (recall)	0.0463 (recall)	0.0 (recall)	0.0463 (recall)
	which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left right contexts into	feature engineering in NLP. Using word embedding as the typical	0.0376 (f-1) - Rouge L:	0.0333 (f-1) - Rouge L:	0.0455 (f-1) - Rouge L:	0.0552 (f-1) - Rouge L:	0.0 (f-1) - Rouge L:	0.0476 (f-1) - Rouge L:
	a fixed-size representation. Afterwards, a simple feedforward neural network	feature representation for words,	0.1132 (prec)	0.0977 (prec)	0.1045 (prec)	0.1757 (prec)	0.1957 (prec)	0.1942 (prec)
	is used to reject or predict entity label for each individual fragment. The	NNs become competitive to	0.1132 (prec) 0.1651 (recall)	0.1193 (recall)	0.1927 (recall)	0.1757 (prec) 0.1193 (recall)	0.1957 (prec) 0.0826 (recall)	0.1942 (prec) 0.1835 (recall)
	proposed method has been evaluated in several popular NER and mention	traditional approaches in NER.	0.1343 (f-1)	0.1193 (fecall) 0.1074 (f-1)	0.1355 (f-1)	0.1421 (f-1)	0.1161 (f-1)	0.1887 (f-1)
	detection tasks, including the CoNLL 2003 NER task and TAC-KBP2015 and TAC-	Many NLP tasks, such as NER,	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	KBP2016 Tri-lingual Entity Discovery and Linking (EDL) tasks. Our methods	chunking and part-of-speech	In this paper, we study a novel	in this paper, we study a novel	In this paper, we study a novel	In @cite, the authors propose a	The Skip-gram model @cite	In @cite, the authors proposed a
	have yielded pretty strong performance in all of these examined tasks. This	(POS) tagging can be formulated	approach for named entity	approach for named entity	approach for named entity	unified neural network	@cite is an efficient method for	unified neural network
	local detection approach has shown many advantages over the traditional	as sequence labeling tasks. In	recognition (NER) and mention	recognition and mention	recognition (NER) and mention	architecture and learning	learning high-quality distributed	architecture and learning
	sequence labelling methods.	@cite , deep convolutional neural	detection in natural language	detection in natural language	detection in natural language	algorithm that can be applied to	vector representations that	algorithm that can be applied to
	(2):	networks (CNN) and conditional	processing. Instead of treating	processing. instead of treating	processing. Instead of treating	various natural language	capture a large number of	various natural language
	We propose a unified neural network architecture and learning algorithm that	random fields (CRF) are used to	NER as a sequence labelling	named entity recognition as a	NER as a sequence labelling	processing tasks including: part-	precise syntactic and semantic	processing tasks including part-
	can be applied to various natural language processing tasks including: part-of-	infer NER labels at a sentence	problem, we propose a new	sequence labelling problem, we	problem, we propose a new	of-speech tagging, chunking,	word relationships. By	of-speech tagging, chunking,
	speech tagging, chunking, named entity recognition, and semantic role	level, where they still use many	local detection approach, which	propose a new local detection	local detection approach, which	named entity recognition, and	subsampling of the frequent	named entity recognition, and
	labeling. This versatility is achieved by trying to avoid task-specific engineering	hand-crafted features to improve	rely on the recent fixed-size	approach, which rely on the	rely on the recent fixed-size	semantic role labeling. In this	words, it obtain significant	semantic role labeling. However,
1	and therefore disregarding a lot of prior knowledge. Instead of exploiting man-	performance, such as	ordinally forgetting encoding	recent fixed-size ordinally	ordinally forgetting encoding	work, we propose a new local	speedup and also learn more	unlike our work, they do not use
	made input features carefully optimized for each task, our system learns	capitalization features explicitly	(FOFE) method to fully encode	forgetting encoding (FOFE )	(FOFE) method to fully encoding	detection approach, which rely	regular word representations.	a deep neural network to train
	internal representations on the basis of vast amounts of mostly unlabeled	defined based on first-letter	each sentence fragment and its	method to fully encode each	each sentence fragment and its	on the recent fixed-size ordinally	Comment:	their system. Instead, they use a
	training data. This work is then used as a basis for building a freely available tagging system with good performance and minimal computational	capital, non-initial capital and so on.	left right contexts into a fixed-	sentence fragment and its left right contexts into a fixed-size	left right contexts into a fixed-	forgetting encoding (FOFE)	<ul> <li>Shows weak signs of MDS, providing a summary of the 3<sup>rd</sup></li> </ul>	multi-task learning algorithm to
	requirements.	on.	size representation. Afterwards, a simple feedforward neural	right contexts into a fixed-size representation.	size representation. Afterwards, a simple feedforward neural	method to fully encode each sentence fragment and its left	article only (blue) and	train their network on a large amount of unlabeled training
	(3):		network is used to reject or	afterwards, a simple	network is used to reject or	right contexts into a fixed-size	presenting it as different from	data. In contrast, we use a deep
	The recently introduced continuous Skip-gram model is an efficient method for		predict entity label for each	feedforward neural network is	predict entity label for each	representation.	the main one.	network to train our system on a
	learning high-quality distributed vector representations that capture a large		individual fragment. The	used to reject or predict entity	individual fragment. The	Comment:	the main one.	much larger amount of training
	number of precise syntactic and semantic word relationships. In this paper we		proposed method has been	label for each individual	proposed method has been	- Shows signs of MDS, which		data, and we do not need to
	present several extensions that improve both the quality of the vectors and		evaluated in several popular NER	fragment.	evaluated in several popular NER	contrasts the 2nd (green) article		train our network on unlabeled
	the training speed. By subsampling of the frequent words we obtain significant		and mention detection tasks,	the proposed method has been	and mention detection tasks,	with the 1st (yellow) one.		data.
	speedup and also learn more regular word representations. We also describe a		including the CoNLL 2003 NER	evaluated in several popular	including the CoNLL 2003 NER	- However, the 3rd article is not		Comment:
	simple alternative to the hierarchical softmax called negative sampling. An		task and TAC-KBP2016 Tri-lingual	named entity and mention	task and TAC-KBP2015 and TAC	covered at all		- Shows signs of MDS, providing
	inherent limitation of word representations is their indifference to word order		Entity Discovery and Linking	detection tasks, including the	-KBP2016 Tri-lingual Entity			a summary of the 2nd article
	and their inability to represent idiomatic phrases. For example, the meanings		(EDL) tasks. Our methods have	coNLL 2003 named entity	Discovery and Linking (EDL)			(green) and contrasting it with
	of "Canada" and "Air" cannot be easily combined to obtain "Air Canada".		yielded pretty strong	recognition task and the	tasks. Our methods have yielded			the main one (red).
	Motivated by this example, we present a simple method for finding phrases in		performance in all of these	tac-2015 and tac-2016 tri-lingual	pretty strong performance in all			- However, the red parts also
	text, and show that learning good vector representations for millions of		examined tasks. This local	entity discovery and linking	of these examined tasks. This			contain hallucinations, not to
	phrases is possible.		detection approach has shown	tasks.	local detection approach has			mention the neglect of the 3 <sup>rd</sup>
			many advantages over the	our methods have yielded	shown many advantages over			article
			traditional sequence labelling	pretty strong performance in all of these examined tasks.	the traditional sequence			
			methods.     We propose a unified neural network	Comment:	labelling methods. We propose a unified neural network			
			architecture and learning	- Copied the first tokens only,	architecture and learning			
			algorithm that can be applied to	except for the deletion of the				
			various natural language	acronym for NER (red)	various natural language			
			processing	- No indication of MDS	processing tasks including: part-			
			Comment:		of-speech tagging, chunking,			
			- Copied the first tokens only		named entity recognition, and			
			- No indication of MDS		semantic role labeling. This			
					versatility is achieved by trying			
					to avoid task-specific			
					engineering and therefore			
					disregarding a lot of prior			
					knowledge. Instead of exploiting			
					man-made input			
					Comment:			
1					Copied the first tokens only     No indication of MDS			
					- NO INDICATION OF MIDS			

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
3524	(1):	More recently, a technique @cite	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	The explosion in the availability of GPS-enabled devices has resulted in an	was designed to study the	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	abundance of trajectory data. In reality, however, majority of these trajectories	trajectory inference problem in a	0.0321 (prec)	0.034 (prec)	0.0051 (prec)	0.0 (prec)	0.0 (prec)	0.0233 (prec)
	are collected at a low sampling rate and only provide partial observations on	setting where trajectories are not	0.1316 (recall)	0.1316 (recall)	0.0263 (recall)	0.0 (recall)	0.0 (recall)	0.0526 (recall)
	their actually traversed routes. Consequently, they are mired with uncertainty.	constrained by a network. Due to	0.0515 (f-1)	0.0541 (f-1)	0.0085 (f-1)	0.0 (f-1)	0.0 (f-1)	0.0323 (f-1)
	In this paper, we develop a technique called Infer Tra to infer uncertain	the focus on network-free	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	trajectories from network-constrained partial observations. Rather than	trajectories, @cite is not	0.0955 (prec)	0.1014 (prec)	0.0653 (prec)	0.2 (prec)	0.1591 (prec)	0.1149 (prec)
	predicting the most likely route, the inferred uncertain trajectory takes the	applicable to our problem.	0.3846 (recall)	0.3846 (recall)	0.3333 (recall)	0.2051 (recall)	0.1795 (recall)	0.2564 (recall)
	form of an edge-weighted graph and summarizes all probable routes in a		0.1531 (f-1)	0.1604 (f-1)	0.1092 (f-1)	0.2025 (f-1)	0.1687 (f-1)	0.1587 (f-1)
	holistic manner. For trajectory inference, Infer Tra employs Gibbs sampling by		Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	learning a Network Mobility Model (NMM) from a database of historical		The explosion in the availability	the explosion in the availability	The world's first GPS-enabled	@cite proposed a routing	@cite proposed a routing	@cite proposed a routing
	trajectories. Extensive experiments on real trajectory databases show that the		of GPS-enabled devices has	of GPS-enabled devices has	vehicle has been around for	algorithm to construct the top-k	algorithm to construct the top-	algorithm to construct the top-
	graph-based approach of Infer Tra is up to 50 more accurate, 20 times faster,		resulted in an abundance of	resulted in an abundance of	more than a decade, but it's not	routes according to a user-	@math routes according to a	@math routes according to a
	and immensely more versatile than state-of-the-art techniques.		trajectory data. In reality,	trajectory data. in reality,	the first time we've seen the use	specified query. The routing	user-specified query. They have	user-specified query. However,
	(2):		however, majority of these	however, majority of these	of GPS-enabled vehicles. In fact,	algorithm is based on	conducted extensive	the routing algorithm does not
	The advances in location-acquisition technologies have led to a myriad of		trajectories are collected at a	trajectories are collected at a	we're only a few years away	collaborative learning among	experiments on two real	consider the spatial and
	spatial trajectories. These trajectories are usually generated at a low or an		low sampling rate and only	low sampling rate and only	from the start of a new	uncertain trajectories and	datasets, namely Foursquare	temporal characteristics of
1	irregular frequency due to applications' characteristics or energy saving,	1	provide partial observations on	provide partial observations on	generation of GPS-powered	construct a routable graph by	check-in datasets and taxi	uncertain trajectories. In
1	leaving the routes between two consecutive points of a single trajectory	1	their actually traversed routes.	their actually traversed routes.	vehicles, which will soon be able	collaborative learning among the	trajectories. The results show	contrast, our routing algorithm is
	uncertain (called an uncertain trajectory). In this paper, we present a Route		Consequently, they are mired	Consequently, they are mired	to connect us to GPS-enabled	uncertain trajectories.	that RICK is both effective and	able to construct the popular
	Inference framework based on Collective Knowledge (abbreviated as RICK) to		with uncertainty. In this paper,	with uncertainty. in this paper,	devices. That means we'll be	Comment:	efficient.	routes from uncertain
	construct the popular routes from uncertain trajectories. Explicitly, given a		we develop a technique called	we develop a technique called	able to use GPS-enabled	<ul> <li>Shows weak signs of MDS,</li> </ul>	Comment:	trajectories in a mutual
	location sequence and a time span, the RICK is able to construct the top-k		Infer Tra to infer uncertain	Infer Tra to infer uncertain	smartphones to track us down	which presents the 2nd (green)	- Shows weak signs of MDS,	reinforcement way (i.e.,
	routes which sequentially pass through the locations within the specified time		trajectories from network-	trajectories from network-	and track us down. But before	article as different from the	which presents the 2nd (green)	uncertain + uncertain →
	span, by aggregating such uncertain trajectories in a mutual reinforcement		constrained partial observations.	constrained partial observations.	we do that, let's take a closer	main one.	article as different from the	certain). Moreover, we do not
	way (i.e., uncertain + uncertain → certain). Our work can benefit trip planning,		Rather than predicting the most	rather than predicting the most	look at what's going on in the	<ul> <li>Red part shows some repeat of</li> </ul>	main one.	consider the temporal and
	traffic management, and animal movement studies. The RICK comprises two		likely route, the inferred	likely route, the inferred	world of GPS-based vehicles.	the previous meaning		spatial characteristics of the
	components: routable graph construction and route inference. First, we		uncertain trajectory takes the	uncertain trajectory takes the	We're going to start with GPS-			uncertain trajectories and
	explore the spatial and temporal characteristics of uncertain trajectories and		form of an edge-weighted graph	form of an edge-weighted graph	enabled cars, which have been			construct a routing algorithm
	construct a routable graph by collaborative learning among the uncertain		and summarizes all probable	and summarizes all probable	around for a long time, but			that can construct the routes
	trajectories. Second, in light of the routable graph, we propose a routing		routes in a holistic manner. For	routes in a holistic manner. for	they're often generated at a low			according to the user-specified
	algorithm to construct the top-k routes according to a user-specified query.		trajectory inference, Infer Tra	trajectory inference,	or irregular frequency due to			<mark>queries.</mark>
	We have conducted extensive experiments on two real datasets, consisting of Foursquare check-in datasets and taxi trajectories. The results show that RICK		employs Gibbs sampling by	we employ Gibbs sampling by	applications' characteristics or			Comment:
	is both effective and efficient.		learning a Network Mobility Model (NMM) from a database	learning a network mobility model ( nMM ) from a database	energy saving, leaving the routes			<ul> <li>Shows signs of MDS, providing a summary of the 2<sup>nd</sup> article</li> </ul>
	is both effective and efficient.		of historical trajectories.	of historical trajectories.	between two consecutive points			(green) and contrasting it with
			Extensive experiments on real	extensive experiments on real	of a single trajectory uncertain (called an uncertain trajectory).			the main one (red).
			trajectory databases show that	trajectory databases show that	In this paper, we present a Route			- However, the red parts also
			the graph-based approach of	the graph-based approach of	Inference framework based on			contain hallucinations, not to
			Infer Tra is up to 50 more	Infer Tra is up to 50 more	Collective Knowledge			mention the neglect of the 3rd
			accurate, 20 times faster, and	accurate, 20 times faster, and	(abbreviated as RICK) to			article
			immensely more versatile than	immensely more versatile than	construct the popular routes			ui dele
			state-of-the-art	state-of-the-art techniques.	from uncertain trajectories.			
			techniques.    The advances	Comment:	Explicitly, given a location			
			in location-acquisition	- Copied the first tokens only,	sequence and a time span, the			
1		1	technologies have led	except for a random	RICK is able to construct the top-			
1		1	Comment:	replacement subject noun	k routes			
		1	- Copied the first tokens only	(red)	Comment:			
		1	- No indication of MDS	- No indication of MDS	- Hallucinated a random			
		1			background on GPS (red)			
		1			before starting to summarize			
		1			the 2 <sup>nd</sup> article (green)			
		1			- No indication of MDS			
		1						
		1	l				l .	

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
	*******					1 11		
4160	(a):  We present a method for extracting depth information from a rectified image pair. We train a convolutional neural network to predict how well two image batthers match and use it to compute the stereo matching cost. The cost is refined by cross-based cost aggregation and semiglobal matching, followed by cross-based cost aggregation and semiglobal matching, followed by elect-right consistency check to eliminate errors in the occuleder regions. Our stereo method achieves an error rate of 2.61 on the KITTI stereo dataset and is currently (August 2014) the top performing method on this dataset.  (2):  While machine learning has been instrumental to the ongoing progress in most areas of computer vision, it has not been applied to the problem of stereo matching with similar frequency or success. Why present a supervised seatons approach for areacting the correctness of stereo matching and as set of feature.  Seaton for the standard of the standard of the standard bread and as est of feature.  Seaton for the standard of the standard of the standard of the standards and in confidence estimation, which allows us to rank pixel according to the reliability of their assigned disparities. Moreover, we show how these confidence values can be used to improve the accuracy of disparity maps by integrating them with an MRF-based stereo algorithm. This is a standard of the standard	Recent work @cite @cite focused on estimating the confidence of the computed matching cost. used a random forest classifier to combine several confidence measures. Similarly, trained a random forest classifier to predict the confidence of the matching cost and used the predictions as soft constraints in a Markov random field to decrease the error of the stereo method.	Rouge scores: - Rouge 2: 0.0532 (prec) 0.0532 (prec) 0.0538 (prec) 0.0578 (fs1) - Rouge 1: 0.0788 (fs1) - Rouge 1: 0.0988 (prec) 0.2931 (recall) 0.1472 (fs1) Summary: We present a method for extracting depth information from a rectified image pair. We train a convolutional neural network to predict how well two image patches match and use it to compute the stereo matching cost. The cost is refined by pross-based cost aggregation and semiglobal matching, followed by a left-right consistency reach to eliminate errors in the competition of the contraction of the competition of the contraction of the competition of the contraction of th	Rouge scores: -Rouge 2: 0.0826 (prec) 0.1579 (recall) 0.1584 (f.1) -Rouge 1: 0.2 (prec) 0.3793 (recall) 0.2619 (f.1) Summary:	Rouge scores: - Rouge 2: 0.0468 (prec) 0.1404 (recall) 0.0702 (£1) - Rouge 1: 0.1407 (prec) 0.3103 (recall) 0.1556 (£1) Summary: In this article, we're going to take a look at a new method for predicting the correctness of steroo starches based on all andion forced and a set of features that capture various forms of information about each pixel. We're using the random' idecision forcet framework, which currently poses higher challenges to steroes solvers than there harchmarks with ground truth for steroe or selection of truth for steroe or selection or septiment with semi global matching steroe or selection or steroe or selection selection selection or selection s	Rouge scores: -Rouge 2: 0.1551 (prec) 0.1556 (prec) 0.1556 (fr) -Rouge 1: 0.2687 (prec) 0.3030 (recall) 0.288 (f-1) Summary: Dut word is dowly related to second a set of features that capture various forms of information about each pixel flowers, and the second control of information about each pixel flowers, and work is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in that we train a convolutional neural network is different in th	Rouge scores: -Rouge 2: 0.0517 (prec) 0.0526 (recall) 0.2337 (prec) 0.2414 (recall) 0.2337 (prec) 0.2414 (recall) 0.2337 (prec) 0.2414 (recall) 0.2337 (rec) 0.2414 (recall) 0.2337 (rec) 0.2414 (recall) 0.2337 (rec) 0.2414 (recall) 0.2414	

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4191	(1):	In @cite , assuming the expected	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Mobile Context-Aware Recommender Systems can be naturally modelled as an	reward of a document is linear,	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	exploration exploitation trade-off (exr exp) problem, where the system has to	they perform recommendation	0.0307 (prec)	0.0625 (prec)	0.0417 (prec)	0.0556 (prec)	0.0743 (prec)	0.0357 (prec)
	choose between maximizing its expected rewards dealing with its current	based on contextual information	0.0926 (recall)	0.1852 (recall)	0.1667 (recall)	0.0556 (recall)	0.2037 (recall)	0.0556 (recall)
	knowledge (exploitation) and learning more about the unknown user's	about the users' documents. To	0.0461 (f-1)	0.0935 (f-1)	0.0667 (f-1)	0.0556 (f-1)	0.1089 (f-1)	0.0435 (f-1)
	preferences to improve its knowledge (exploration). This problem has been	maximize the total number of	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	addressed by the reinforcement learning community but they do not consider	user's clicks, this work proposes	0.0976 (prec)	0.1304 (prec)	0.0737 (prec)	0.1636 (prec)	0.1678 (prec)	0.1294 (prec)
	the risk level of the current user's situation, where it may be dangerous to	the LINUCB algorithm which is	0.2909 (recall)	0.3818 (recall)	0.2909 (recall)	0.1636 (recall)	0.4545 (recall)	0.2 (recall)
	recommend items the user may not desire in her current situation if the risk level is high. We introduce in this paper an algorithm named R-UCB that	computationally efficient if the expected rewards of documents	0.1461 (f-1) Summary:	0.1944 (f-1) Summary:	0.1176 (f-1) Summary:	0.1636 (f-1) Summary:	0.2451 (f-1) Summary:	0.1571 (f-1) Summary:
	considers the risk level of the user's situation to adaptively balance between	are linear which is not always the	Mobile Context-Aware	personalized web services strive	A new algorithm for	In @cite, the authors propose a	In @cite, the authors model	In @cite, the authors propose a
	exr and exp. The detailed analysis of the experimental results reveals several	case.	Recommender Systems can be	to adapt their services	personalized recommendation	contextual bandit algorithm that	personalized recommendation	contextual bandit algorithm that
	important discoveries in the exr exp behaviour.	cusc.	naturally modelled as an	(advertisements, news articles )	of news articles has been	is computationally efficient and	of news articles as a contextual	is computationally efficient and
	(2):		exploration exploitation trade-	to individual users by making	developed by the reinforcement	well motivated from learning	bandit problem, a principled	well motivated from learning
	Personalized web services strive to adapt their services (advertisements, news		off (exr exp) problem, where the	use of both content and user	learning community, but it does	theory. They argue that any	approach in which a learning	theory. They argue that any
	articles, etc.) to individual users by making use of both content and user		system has to choose between	information. despite a few	not consider the risk level of the	bandit algorithm can be reliably	algorithm sequentially selects	algorithm that can reliably
	information. Despite a few recent advances, this problem remains challenging		maximizing its expected rewards	recent advances, this problem	current user's situation, where it	evaluated offline using	articles to serve users based on	evaluate offline using previous
	for at least two reasons. First, web service is featured with dynamically		dealing with its current	remains challenging for at least	may be dangerous to	previously recorded random	contextual information about	recorded random traffic is
	changing pools of content, rendering traditional collaborative filtering		knowledge (exploitation) and	two reasons.	recommend items the user may	traffic. Finally, they successfully	the users and articles, while	reliably evaluated. They also
	methods inapplicable. Second, the scale of most web services of practical		learning more about the	first, web service is featured	not desire in her current	applied their new algorithm to a	simultaneously adapting its	argue that any offline evaluation
	interest calls for solutions that are both fast in learning and computation. In		unknown user's preferences to	with dynamically changing pools	situation if the risk level is high.	Yahoo! Front Page Today Module	article selection strategy based	method can be reliably applied
	this work, we model personalized recommendation of news articles as a contextual bandit problem, a principled approach in which a learning		improve its knowledge (exploration). This problem has	of content, rendering traditional collaborative filtering methods	We introduce in this paper an algorithm named R-UCB that	dataset containing over 33 million events.	on user-click feedback to maximize total user clicks. The	to a large dataset containing over 33 million events. The
	algorithm sequentially selects articles to serve users based on contextual		been addressed by the	inapplicable.	considers the risk level to	Comment:	rontributions of this work are	results showed a 12.5 click lift
	information about the users and articles, while simultaneously adapting its		reinforcement learning	second, the scale of most web	adaptively balance between exr	- Shows weak signs of MDS,	threefold. First, they propose a	compared to a standard context-
	article-selection strategy based on user-click feedback to maximize total user		community but they do not	services of practical interest calls	and exp. The detailed analysis of	which presents the 2 <sup>nd</sup> (green)	new, general contextual bandit	free bandit algorithm, and the
	clicks. The contributions of this work are three-fold. First, we propose a new,		consider the risk level of the	for solutions that are both fast in	the experimental results reveals	article as different from the	algorithm that is	advantage becomes even
	general contextual bandit algorithm that is computationally efficient and well		current user's situation, where it	learning and computation. in	several important discoveries in	main one	computationally efficient and	greater when data gets more
	motivated from learning theory. Second, we argue that any bandit algorithm		may be dangerous to	this work, we model	the exr exp behaviour.		well motivated from learning	scarce.
	can be reliably evaluated offline using previously recorded random traffic.		recommend items the user may	personalized recommendation	Personalized web services strive		theory. Second, they argue that	Comment:
	Finally, using this offline evaluation method, we successfully applied our new		not desire in her current	of news articles as a contextual	to adapt their services		any bandit algorithm can be	-Shows weak signs of MDS,
	algorithm to a Yahoo! Front Page Today Module dataset containing over 33		situation if the risk level is high.	bandit problem, a principled	(advertisements, news articles,		reliably evaluated offline using	which presents the 2nd (green)
	million events. Results showed a 12.5 click lift compared to a standard context-		We introduce in this paper an	approach in which a learning	etc.) to individual users by		previously recorded random	article as different from the
	free bandit algorithm, and the advantage becomes even greater when data		algorithm named R-UCB that	algorithm sequentially selects	making use of both content and		traffic. Finally, using this offline	main one
	gets more scarce.		considers the risk Level of the user's situation to adaptively	articles to serve users based on contextual information about	user information. Despite a few recent advances, this problem		evaluation method, they	<ul> <li>The 1st red part contains weird phrasing while the 2nd one is a</li> </ul>
			balance between exr and exp.	the users and articles, while	recent advances, this problem remains challenging for at least		successfully applied their new algorithm to a Yahoo! Front Page	hallucination
			The detailed analysis of the	simultaneously adapting its	two reasons. First, web service is		Today Module dataset	Halidelilation
			experimental results reveals	article-selection strategy based	featured with dynamically		containing over 33 million	
			several important discoveries in	on user-click feedback to	changing pools of content,		events. Results showed a 12.5	
			the exr exp behaviour.	maximize total user clicks.	rendering traditional		click lift compared to a standard	
			Personalized web services strive	we propose a new, general	collaborative filtering methods		context-free bandit algorithm,	
			to adapt their services	contextual bandit algorithm that	inapplicable. Second, the scale		and the advantage becomes	
			(advertisements, news articles,	is computationally efficient and	of most web services of practical		even greater when data gets	
			etc.) to individual users by	well motivated from learning	interest calls for solutions that		more scarce.	
			making use of both content and	theory.	are both fast in learning and		Comment:	
			user information. Despite Comment:	moreover, we argue that any bandit algorithm can be reliably	computation. In this work, we model personalized		<ul> <li>Shows weak signs of MDS, which presents the 2<sup>nd</sup> (green)</li> </ul>	
			- Copied the first tokens only	evaluated offline using	recommendation ofnews articles		article as different from the	
			Copied the first tokens only     No indication of MDS	Comment:	as a contextual bandit problem.		main one	
			- NO INDICACION OF WIDS	- Copied from the 2 <sup>nd</sup> article only	as a contextual bandit problem, a principled approach in which a		-Some creative writing noted	
				- No indication of MDS	learning algorithm sequentially		(red)	
					selects articles to serve users		,	
					based on contextual information			
					about the users and articles,			
					while simultaneously adapting			
					its article-selection strategy			
					based on user-click feedback			
					Comment:			
					-Simply joined together selected			
					parts of the 1st (yellow) and 2nd (green) articles			
					- No indication of MDS			
					- NO INDICATION OF IVIDS			
1	1		1	I	I		I	1

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4820	(1): This paper presents a new method and a constraint-based objective function	The two other methods described in @cite are based on the idea	Rouge scores: - Rouge 2:	Rouge scores: - Rouge 2:	Rouge scores: - Rouge 2:	Rouge scores: - Rouge 2:	Rouge scores: - Rouge 2:	Rouge scores: - Rouge 2:
	to solve two problems related to the design of optical telecommunication	that to save ADMs a good	0.0132 (prec)	0.0 (prec)	0.0205 (prec)	0.0189 (prec)	0.0357 (prec)	0.1233 (prec)
'	networks, namely the Synchronous Optical Net-work Ring Assignment Problem	solution should have very dense	0.0144 (recall)	0.0 (recall)	0.0288 (recall)	0.0072 (recall)	0.0072 (recall)	0.0647 (recall)
	(SRAP) and the Intra-ring Synchronous Optical Network DesignProblem (IDP).	rings. They are both greedy and	0.0138 (f-1)	0.0 (f-1)	0.024 (f-1)	0.0104 (f-1)	0.012 (f-1)	0.0849 (f-1)
	These network topology problems can be represented as a graph partitioning	rely on a clique algorithm. In	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	withcapacity constraints as shown in previous works. W'e present here a new	graph theory, a clique in an	0.1316 (prec)	0.1579 (prec)	0.1276 (prec)	0.2407 (prec)	0.2414 (prec)	0.2703 (prec)
	objective function and anew local search algorithm to solve these problems.	undirected graph @math is a	0.1429 (recall)	0.0857 (recall)	0.1786 (recall)	0.0929 (recall)	0.05 (recall)	0.1429 (recall)
	Experiments conducted in C	subset of the vertex set @math ,	0.137 (f-1)	0.1111 (f-1)	0.1488 (f-1)	0.134 (f-1)	0.0828 (f-1)	0.1869 (f-1)
	(2):	such that for every two vertices in	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	This paper considers two problems that arise in the design of optical	@math , there exists an edge	This paper presents a new	this paper presents a new	This paper presents a new	In the context of optical	In @cite, an objective function	In @cite, the objective functions
	telecommunication networks when a ring-based topology is adopted, namely	connecting the two. Finding a	method and a constraint-based	method and a constraint-based	method and a constraint-based	telecommunication networks,	and a diversification method are	of the SONET ring assignment
l.	the SONET Ring Assignment Problem and the Intraring Synchronous Optical	clique is not that easy, a way to	objective function to solve two	objective function to solve two	objective function to solve two	the SONET Ring Assignment	proposed to solve the SONET	problem and the IDP problem
	Network Design Problem. We show that these two network topology	do it is to use an "Union-Find"	problemsrelated to the design of	problemsrelated to the design of	problems related to the design	Problem (SRAP) and the	Ring Assignment Problem (SRAP)	are discussed in terms of a
l.	problems correspond to graph partitioning problems with capacity constraints:	strategie, two clique @math and	optical telecommunication	optical telecommunication	of optical telecommunication	Intraring Synchronous Optical	and the Intra-ring Synchronous	metaheuristic algorithm. In
l.	the first is a vertex partitioning problem, while the latter is an edge partitioning problem. We consider solution methods for both problems, based	@math such that each node in @math is adjacent to each node	networks, namely the	networks, namely the Synchronous Optical Net-work	networks, namely the Synchronous Optical Net-work	Network Design Problem (IDP) are two problems that arise	Optical Network Design Problem (IDP).	particular, the objective function for SONET Ring Assignment
			Synchronous Optical Net-work				Comment:	Problem is described in @cite.
	on metaheuristic algorithms. We first describe variable objective functions that depend on the transition from one solution to a neighboring one, then we	in @math then merge the two cliques (). The associated	Ring Assignment Problem (SRAP) and the Intra-ring Synchronous	Ring Assignment Problem (SRAP) and the Intra-ring	Ring Assignment Problem (SRAP) and the Intraring Synchronous	when a ring-based topology is adopted. The SRAP problem can	-Shows weak signs of MDS,	The objective functions of SONet
	apply several diversification and intensification techniques including Path	heuristic starts by considering	optical Network DesignProblem	Synchronous optical Network	optical Network DesignProblem	be represented as a graph	which presents the 2nd (green)	Ring Assignment Problem are
	Relinking, exploring Tabu Search and Scatter Search. Finally we propose a	each node to be a clique of size	(IDP). These network topology	DesignProblem (IDp.).	(IDP). These network topology	partitioning problem, while the	article as different from the	defined as: where @math is the
l.	diversification method based on the use of multiple neighborhoods. A set of	one, and to merge two cliques	problems can be represented as	these network topology	problems can be represented as	IDP problem is a edge	main one	number of neighbors, @math
Į.	extensive computational results is used to compare the behaviour of the	into a larger clique until there are	a graph partitioning	problems can be represented as	a graph partitioning	partitioning problem.	main one	the number of nodes in the
l	proposed methods and objective functions.	no more possible merges.	withcapacity constraints as	a graph partitioning	withcapacity constraints as	Comment:		network, @math and @math
	1		shown in previous works. We	withcapacity constraints as	shown in previous works. We	- Shows no sign of MDS, with		are the number of vertices in the
Į.	ĺ		present here a new objective	shown in previous works.	present here a new objective	the model just extracting the		network.
	ĺ		function and anew local search	we present here a new	function and anew local search	common theme of the 2		Comment:
	ĺ		algorithm to solve these	objective function and anew	algorithm to solve these	articles		-Shows weak signs of MDS,
	ĺ		problems. Experiments	local search algorithm to solve	problems. Experiments			which presents the 2 <sup>nd</sup> (green)
l.	ĺ		conducted in CERN and the	these problems.	conducted in C This paper			article as different from the
	ĺ		University of California,	Comment:	considers two problems that			main one
	ĺ		Berkeley.   This paper considers	- Copied the first tokens only	arise in the design of Optical			-Contains more relevant
l.	ĺ		two problems that arise in the	- No indication of MDS	telecommunication networks			information that the other
	ĺ		design of Optical		when a ring-based topology is			tuned models
	ĺ		telecommunication networks		adopted, namely the SONET			l .
l.	ĺ		when a ring-based topology is adopted, namely the SONET		Ring Assignment Problem and the Intra-ring			1
	ĺ		Ring Assignment Problem and		SynchronousOptical Network			l .
	ĺ		the Intraring Synchronous O		Design Problem. We show that			1
l.	ĺ		Network Design Problem. We		these two network topology			1
	ĺ		show that these two network		problem correspond to graph			l .
l.	ĺ		topologies problems correspond		partitioning problems with			l .
Į.	ĺ		to graph partitioning problems		capacity constraints: the first is a			l .
Į.	ĺ		with capacity constraints: the		partitioning problem, while the			l .
	ĺ		first is a vertex partitioning		latter is an edge partitioning			1
	ĺ		problem, while the latter is an		problem. We consider solution			l .
	ĺ		edge partitioning problem.		methods for both problems,			1
	ĺ		Comment:		based on metaheuristic			l .
	ĺ		- Copied the first tokens only		algorithms. We first describe			l .
	ĺ		- No indication of MDS		variable objective functions that			1
	ĺ				depend on the transition from			1
	ĺ				one solution to a neighboring			1
Į.	1				one, then we apply several diversification and			l
	1				intensification techniques			l
	1				including Path Relinking,			l
	1				eXploring Tabu Search and			l
ı	1				Scatter Search. Finally we			
ı	1				propose a divers			
ı	1				Comment:			1
	1				- Copied the first tokens only			l
l.	1				- No indication of MDS			1
l	1							1
	<u> </u>							

(b) Medium samples

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
235	(1):	Several recent papers have	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	We present a data-driven framework for incorporating side information in	focused on tractable	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	dynamic optimization under uncertainty. Specifically, our approach uses	approximations of two- and multi-	0.0812 (prec)	0.0917 (prec)	0.1179 (prec)	0.2093 (prec)	0.2708 (prec)	0.046 (prec)
	predictive machine learning methods (such as k-nearest neighbors, kernel	stage and robust optimization.	0.0823 (recall)	0.0696 (recall)	0.1582 (recall)	0.057 (recall)	0.0823 (recall)	0.0253 (recall)
	regression, and random forests) to weight the relative importance of various data-driven uncertainty sets in a robust optimization formulation. Through a	Many approaches are based around policy approximation	0.0818 (f-1) - Rouge L:	0.0791 (f-1) - Rouge L:	0.1351 (f-1) - Rouge L:	0.0896 (f-1) - Rouge L:	0.1262 (f-1) - Rouge L:	0.0327 (f-1) - Rouge L:
	novel measure concentration result for local machine learning methods, we	schemes, including lifted linear	0.1553 (prec)	0.1983 (prec)	0.1784 (prec)	0.3409 (prec)	0.4082 (prec)	0.1932 (prec)
	prove that the proposed framework is asymptotically optimal for stochastic	decision rules , @math -adaptivity	0.1572 (recall)	0.1509 (recall)	0.239 (recall)	0.0943 (recall)	0.4062 (prec) 0.1258 (recall)	0.1069 (recall)
	dynamic optimization with covariates. We also describe a general-purpose	, and finite adaptability .	0.1562 (f-1)	0.1714 (f-1)	0.2043 (f-1)	0.1478 (f-1)	0.1923 (f-1)	0.1377 (f-1)
	approximation for the proposed framework, based on overlapping linear	Alternative approaches include	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	decision rules, which is computationally tractable and produces high-quality	tractable approximations of	We present a data-driven	we present a data-driven	A new approach to stochastic	In @cite, the authors propose a	In the context of stochastic	In @cite @cite, the authors
	solutions for dynamic problems with many stages. Across a variety of examples	copositive formulations . Closest	framework for incorporating side	framework for incorporating side	dynamic optimization has been	linear decision rule for each	dynamic optimization with	propose a framework for
	in shipment planning, inventory management, and finance, our method	related to the approximation	information in dynamic	information in dynamic	developed by a team of	uncertainty set. However, their	covariates, @cite and @cite	stochastic dynamic optimization,
	achieves improvements of up to 15 over alternatives and requires less than one	scheme in this paper are @cite	optimization under uncertainty.	optimization under uncertainty.	researchers from the University	approach is not as robust as our	proposed methods for optimizing	which is computationally
	minute of computation time on problems with twelve stages.	and @cite , which address two-	Specifically, our approach uses	our approach uses predictive	of California, Los Angeles, and	approach, which is based on	a separate linear decision rule for	tractable and efficient. However,
	(2): We investigate a data-driven approach to two-stage stochastic linear	stage problems via overlapping decision rules. @cite propose a	predictive machine learning methods (such as k-nearest	machine learning methods ( such as k-nearest neighbors, kernel	the University of Southern California. The new approach is	overlapping linear decision rules. In contrast, our approach is	each uncertainty set. In @cite, the authors proposed a lifting	their framework is not asymptotically asymptotic, since
	optimization in which an uncertainty set is constructed around each data point.	modeling approach that leads to	neighbors, kernel regression, and	regression, and random forests )	based on the use of predictive	asymptotically optimal for	technique that maps a given	it does not require any additional
	We propose an approximation algorithm for these sample robust optimization	novel approximations of various	random forests) to weight the	to weight the relative	machine learning methods (such	stochastic dynamic optimization	stochastic program to an	stages, and we do not need any
	problems by optimizing a separate linear decision rule for each uncertainty set.	distributionally robust	relative importance of various	importance of various data-	as k-nearest neighbors, kernel	with covariates.	equivalent problem on a higher-	additional stages. Moreover, the
	We show that the proposed algorithm combines the asymptotic optimality and	applications, including two-stage	data-driven uncertainty sets in a	driven uncertainty sets in a	regression, and random forests)	Comment:	dimensional probability space.	authors in @cite show that their
	scalability of the sample average approximation while simultaneously offering	distributionally robust	robust optimization formulation.	robust optimization formulation.	to weight the relative	- Shows sign of MDS, with the	Comment:	framework is computationally
	improved out-of-sample performance guarantees. The practical value of our	optimization using Wasserstein	Through a novel measure	through a novel measure	importance of various data-	sentence in red contrasting the	- Clear sign of MDS, with the	optimal for the stochastic
	method is demonstrated in network inventory management and hospital	ambiguity sets and expectations	concentration result for local	concentration result for local	driven uncertainty sets in a	1st (yellow) and 2nd articles.	sentence context of the main	setting, which is more tractable
	scheduling. (3):	of piecewise convex objective functions in single-stage	machine learning methods, we prove that the proposed	machine learning methods , we prove that the proposed	robust optimization formulation. Through a novel measure	<ul> <li>However, the model failed to mention anything about the 3<sup>rd</sup></li> </ul>	article given (yellow) and the 2 <sup>nd</sup> (green) and 3 <sup>rd</sup> (blue)	than our method. Moreover, our
	Stochastic programming provides a versatile framework for decision-making	problems. Independently, @cite	framework is asymptotically	framework is asymptotically	concentration result for local	article	articles clearly mentioned.	dynamic problems in which
	under uncertainty, but the resulting optimization problems can be	investigate a of two-stage sample	optimal for stochastic dynamic	optimal for stochastic dynamic	machine learning methods, we	UI DETE	articles clearly mentioned.	multiple stages are required,
	computationally demanding. It has recently been shown that primal and dual	robust optimization by optimizing	optimization with covariates. We	optimization with covariates.	prove that the proposed			which is not the case in our case.
	linear decision rule approximations can yield tractable upper and lower bounds	a separate linear decision rule for	also describe a general-purpose	we also describe a general-	framework is asymptotically			Comment:
	on the optimal value of a stochastic program. Unfortunately, linear decision	each uncertainty set and prove	approximation for the proposed	purpose approximation for the	optimal for stochastic Dynamic			- Some signs of MDS, with the
	rules often provide crude approximations that result in loose bounds. To	that this approximation gap	framework, based on	proposed framework, based on	Optimization with stochastic			red parts attempting to make
	address this problem, we propose a lifting technique that maps a given	converges to zero as the amount	overlapping linear decision rules,	overlapping linear decision rules,	optimization. We also describe a			contrasts.
	stochastic program to an equivalent problem on a higher-dimensional probability space. We prove that solving the lifted problem in primal and dual	of data goes to infinity. In of this paper, we show how to extend	which is computationally tractable and produces high-	which is computationally tractable and produces high-	general-purpose approximation for the proposed framework,			<ul> <li>However, the model wrongly uses words of the main article</li> </ul>
	linear decision rules provides tighter bounds than those obtained from applying	similar techniques to dynamic	quality solutions for dynamic	quality solutions for dynamic	based on overlapping linear			(vellow) to describe the 2 <sup>nd</sup> and
	linear decision rules provides tighter bounds than those obtained from applying linear decision rules to the original problem. We also show that there is a one-	problems with many stages for	problems with many stages.	problems with many stages.	decision rules, which is			3rd articles.
	to-one correspondence between linear decision rules in the lifted problem and	the first time.	Across a variety of examples in	the practical value of our	computationally tractable and			
	families of nonlinear decision rules in the original problem. Finally, we identify		shipment planning, inventory	method is demonstrated in	produces high-quality solutions			
	structured liftings that give rise to highly flexible piecewise linear and nonlinear		management, and finance, our	network inventory management	for dynamic problems with many			
	decision rules, and we assess their performance in the context of a dynamic		method achieves improvements	and hospital scheduling.	stages. Across a variety of			
	production planning problem.		of up to 15 over alternatives and	Comment:	examples in shipment planning,			
			requires less than one minute of computation time on problems	<ul> <li>Mostly copied the first tokens, except for the dropping of the</li> </ul>	inventory management, and finance, our method achieves			
			with twelve stages.    We	word "Specifically" and the	improvements of up to 15 over			
			investigate a data-based	replacement of the last	alternatives and requires less			
			approach to two-stage stochastic	sentence with the green part	than one minute of computation			
			linear optimization in which an	- No indication of MDS	time on problems with twelve			
			uncertainty set is		stages. We investigate a data-			
			Comment:		driven approach to two-stage			
			- Copied the first tokens only		stochastic linear optimization in			
			- No indication of MDS		which an uncertainty set is			
1					constructed around each data point. We propose an			
					approximation algorithm for			
1					these sample robust			
1					optimization problems by			
					optimizing a separate linear			
					decision rule for each			
					uncertainty set. We show that			
					the proposed algorithm  Comment:			
1					- Except for the randomly			
					hallucinated beginning (red),			
					the model just copied the 1st			
					article in full, followed by the			
					initial parts of the 2 <sup>nd</sup> .			ļ ļ
-	•							

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
638	(1):	Concurrently with our work, some	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Previous works utilized "smaller-norm-less-important" criterion to prune filters	data independent filter pruning	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	with smaller norm values in a convolutional neural network. In this paper, we	strategies @cite @cite @cite	0.0483 (prec)	0.037 (prec)	0.0193 (prec)	0.0658 (prec)	0.0506 (prec)	0.0412 (prec)
	analyze this norm-based criterion and point out that its effectiveness depends	@cite have been explored. @cite	0.0986 (recall)	0.0845 (recall)	0.0563 (recall)	0.0704 (recall)	0.1127 (recall)	0.0563 (recall)
	on two requirements that are not always met: (1) the norm deviation of the	utilizes an @math -norm criterion	0.0648 (f-1)	0.0515 (f-1)	0.0288 (f-1)	0.068 (f-1)	0.0699 (f-1)	0.0476 (f-1)
	filters should be large; (2) the minimum norm of the filters should be small. To	to prune unimportant filters.	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	solve this problem, we propose a novel filter pruning method, namely Filter	@cite proposes to select filters	0.1301 (prec)	0.1043 (prec)	0.0769 (prec)	0.1558 (prec)	0.0943 (prec)	0.1327 (prec)
	Pruning via Geometric Median (FPGM), to compress the model regardless of	with an @math -norm criterion	0.2639 (recall)	0.2361 (recall)	0.2222 (recall)	0.1667 (recall)	0.2083 (recall)	0.1806 (recall)
	those two requirements. Unlike previous methods, FPGM compresses CNN models by pruning filters with redundancy, rather than those with "relatively	and prune those selected filters in a soft manner. @cite proposes to	0.1743 (f-1) Summary:	0.1447 (f-1) Summary:	0.1143 (f-1) Summary:	0.1611 (f-1) Summary:	0.1299 (f-1) Summary:	0.1529 (f-1) Summary:
	less" importance. When applied to two image classification benchmarks, our	prune models by enforcing	Previous works utilized "smaller-	model pruning has become a	Deep Convolutional Neural	In <u>@cite</u> , the <u>authors proposed a</u>	@cite proposed a filter pruning	In @cite, the authors proposed
	method validates its usefulness and strengths. Notably, on CIFAR-10, FPGM	sparsity on the scaling parameter	norm-less-important" criterion	useful technique that improves	Networks (CNN) has achieved	filter pruning method to	method based on spectral	spectral clustering filter pruning
	reduces more than 52 FLOPs on ResNet-110 with even 2.69 relative accuracy	of batch normalization layers.	to prune filters with smaller	the computational efficiency of	significant success in computer	compress CNN models by	clustering filter pruning with soft	with soft self-adaption manners
	improvement. Moreover, on ILSVRC-2012, FPGM reduces more than 42 FLOPs	@cite uses spectral clustering on	norm values in a convolutional	deep learning, making it possible	vision field. However, the high	pruning filters with smaller norm	self-adaption manners (SCSP).	(SCSP). They first applied spectral
	on ResNet-101 without top-5 accuracy drop, which has advanced the state-of-	filters to select unimportant ones.	neural network. In this paper, we	to deploy solutions on resource-	computational cost of the deep	values. However, their method	They first apply spectral	clustering on filters layer by layer
	the-art. Code is publicly available on GitHub: this https URL		analyze this norm-based	limited scenarios. in this paper,	complex models prevents the	does not require the support of	clustering on filters layer by layer	to explore their intrinsic
	(2):		criterion and point out that its	we propose a channel pruning	deployment on edge devices	sparse convolution libraries.	to explore their intrinsic	connections and only count on
	Deep Convolutional Neural Networks (CNN) has achieved significant success in		effectiveness depends on two	technique for accelerating the	with limited memory and	Moreover, they do not need the	connections and only count on	efficient groups. Then, the
	computer vision field. However, the high computational cost of the deep	1	requirements that are not always	computations of deep	computational resource. In this	support of CNN libraries and can	efficient groups. @cite proposed	network gradually selects
	complex models prevents the deployment on edge devices with limited memory	1	met: (1) the norm deviation of	convolutional neural networks	paper, we propose a novel filter	work with existing efficient BLAS	an approach to accelerate the computations of deep	meaningful groups. The authors
	and computational resource. In this paper, we proposed a novel filter pruning for convolutional neural networks compression, namely spectral clustering filter		the filters should be large; (2) the minimum norm of the filter	(CNNs ), which does not critically rely on the assumption that a	pruning for convolutional neural networks compression, namely	libraries. To solve this problem, they proposed a method called	computations of deep convolutional neural networks	also proposed a novel angle to interpret the model compression
	pruning with soft self-adaption manners (SCSP). We first apply spectral		should be small. To solve this	smaller- norm parameter or	spectral clustering filter pruning	Filter Pruning via Geometric	(CNNs), which does not critically	process. In @cite, the authors
	clustering on filters layer by layer to explore their intrinsic connections and only		problem, we propose a novel	feature plays a less informative	with soft self-adaption manners	Median (FPGM) to prune filters	rely on this assumption. Instead,	proposed a novel method for
	count on efficient groups. By self-adaption manners, the pruning operations can		filter pruning method, namely	role at the inference time.	(SCSP). We first apply spectral	with larger norm values in a	it focuses on direct simplification	convolutional neural networks
	be done in few epochs to let the network gradually choose meaningful groups.		Filter Pruning via Geometric	our approach takes two stages :	clustering on filters layer by layer	convolutional neural network.	of the channel-to-channel	compression, which is based on
	According to this strategy, we not only achieve model compression while		Median (FPGM), to compress the	the first being to adopt an end-	to explore their intrinsic	Comment:	computation graph of a CNN	spectral clustering. They
	keeping considerable performance, but also find a novel angle to interpret the		model regardless of those two	to-end stochastic training	connections and only count on	- Shows sign of MDS, attempting	without the need of performing	proposed a novel algorithm to
	model compression process.		requirements. Unlike previous	method that eventually forces	efficient groups. By self-	to provide a summary of the 4th	a computational difficult and not	interpret the network
	(3):		methods, FPGM compresses	the outputs of some channels	adaptions manners, the pruning	article (purple) and contrast	always useful task of making	compression process, and they
	Model pruning has become a useful technique that improves the computational		CNN models by pruning filters	being constant, and the second	operations can be done in few epochs to let the network	with the main one	high-dimensional tensors of CNN structured sparse. Their	also found a novel angle for
	efficiency of deep learning, making it possible to deploy solutions on resource- limited scenarios. A widely-used practice in relevant work assumes that a		with redundancy, rather than those with "relatively less"	being to prune those constant channels from the original neural	gradually choose meaningful	<ul> <li>Of particular note is that, despite the 1st yellow part</li> </ul>	approach takes two stages: the	interpret the compression
	smaller- norm parameter or feature plays a less informative role at the inference		importance. When applied to	network by adjusting the biases	groups. According to this	came from the 1st article, the	first being to adopt an end-to-	Comment:
	time. In this paper, we propose a channel pruning technique for accelerating the		two image classification	of their impacting layers such	strategy, we not only achieve	model was able to present it as	end stochastic training method	- Shows some sign of MDS.
	computations of deep convolutional neural networks (CNNs), which does not		benchmarks, our method	that the resulting compact model	model compression while	not related to the main article	that eventually forces the	attempting to summarize
	critically rely on this assumption. Instead, it focuses on direct simplification of		validates its usefulness and	can be quickly fine-tuned.	keeping considerable	- However, it made an error by	outputs of some channels being	multiple articles (green and red
	the channel-to-channel computation graph of a CNN without the need of		strengths. FPGM reduces more	we first apply spectral clustering	performance, but also find a	stating FPGM as linked to the	constant, and the second being	@cite)
	performing a computational difficult and not always useful task of making high-		than 52 FLOPs on ResNet-110	on filters layer by layer to explore	novel angle to interpret the	other articles	to prune those constant	- However, content-wise, only
	dimensional tensors of CNN structured sparse. Our approach takes two stages:		without top-5 accuracy drop,	their intrinsic connections and	model compression process. A		channels from the original neural	the 2 <sup>nd</sup> article is covered
	the first being to adopt an end-to-end stochastic training method that		which has advanced the state-of- the-art	only count on efficient groups. by	widely-used practice in relevant		network by adjusting the biases	
	eventually forces the outputs of some channels being constant, and the second being to prune those constant channels from the original neural network by		Comment:	self-adaption manners, the pruning operations can be done	work assumes that a smaller- norm parameter or feature plays		of their impacting layers such that the resulting compact model	
	adjusting the biases of their impacting layers such that the resulting compact		- Copied the first tokens only	in few epochs to	a less informative role at the		can be quickly fine-tuned.	
	model can be quickly fine-tuned. Our approach is mathematically appealing		No indication of MDS	Comment:	inference time. In this article, we		Comment:	
	from an optimization perspective and easy to reproduce. We experimented our			- Extracted sentences from the	propose the channel pruning		- Shows clear sign of MDS,	
	approach through several image learning benchmarks and demonstrate its			3rd article (blue) plus another	technique for accelerating the		attempting to summarize the	1
	interesting aspects and the competitive performance.			from the 2 <sup>nd</sup> (green) only	computations of deep		2 <sup>nd</sup> (green) and 3 <sup>rd</sup> articles	1
	(4):			- No indication of MDS	convolutional Neural networks		(blue)	
	The success of CNNs in various applications is accompanied by a significant				(CNNs), which does not critically		- 4th article is regrettably not	
	increase in the computation and parameter storage costs. Recent efforts toward				rely on this assumption. Instead, it focuses on direct simplification		covered	1
	reducing these overheads involve pruning and compressing the weights of various layers without hurting original accuracy. However, magnitude-based				of the channel-to-channel			
	pruning of weights reduces a significant number of parameters from the fully				computation graph of a CNN			1
	connected layers and may not adequately reduce the computation costs in the				without the need of performing			1
	convolutional layers due to irregular sparsity in the pruned networks. We				a computational difficult and not			1
	present an acceleration method for CNNs, where we prune filters from CNNs				always useful task			1
	that are identified as having a small effect on the output accuracy. By removing				Comment:			
	whole filters in the network together with their connecting feature maps, the				-Copied just the whole 2nd			1
	computation costs are reduced significantly. In contrast to pruning weights, this				article (green), then started			
	approach does not result in sparse connectivity patterns. Hence, it does not				copying from the 2 <sup>nd</sup> sentence of the 3 <sup>rd</sup> article (blue).			
	need the support of sparse convolution libraries and can work with existing efficient BLAS libraries for dense matrix multiplications. We show that even				of the 3rd article (blue).  - No indication of MDS			
	simple filter pruning techniques can reduce inference costs for VGG-16 by up to				- NO INDICATION OF MDS			
	34 and ResNet-110 by up to 38 on CIFAR10 while regaining close to the original							
	accuracy by retraining the networks.							

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
831	(1):	Multi-task learning (MTL) research	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Learning to drive faithfully in highly stochastic urban settings remains an open	shows the joint training of	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	problem. To that end, we propose a Multi-task Learning from Demonstration	auxiliary related side-tasks along	0.0721 (prec)	0.0276 (prec)	0.0192 (prec)	0.0492 (prec)	0.0494 (prec)	0.0541 (prec)
	(MT-LfD) framework which uses supervised auxiliary task prediction to guide the	with the main task enhances the	0.0748 (recall)	0.0374 (recall)	0.0374 (recall)	0.028 (recall)	0.0374 (recall)	0.0561 (recall)
	main task of predicting the driving commands. Our framework involves an end-	training performance . MTL in	0.0734 (f-1)	0.0317 (f-1)	0.0254 (f-1)	0.0357 (f-1)	0.0426 (f-1)	0.055 (f-1)
	to-end trainable network for imitating the expert demonstrator's driving	neural networks has been	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	commands. The network intermediately predicts visual affordances and action primitives through direct supervision which provide the aforementioned	successfully demonstrated in many tasks previously including	0.1786 (prec)	0.1301 (prec) 0.1759 (recall)	0.0861 (prec) 0.1667 (recall)	0.1774 (prec)	0.1585 (prec) 0.1204 (recall)	0.1786 (prec) 0.1852 (recall)
	auxiliary supervised guidance. We demonstrate that such joint learning and	text-to-speech conversion ,	0.1852 (recall) 0.1818 (f-1)	0.1759 (recall) 0.1496 (f-1)	0.1667 (recall) 0.1136 (f-1)	0.1019 (recall) 0.1294 (f-1)	0.1204 (recall) 0.1368 (f-1)	0.1852 (recall) 0.1818 (f-1)
	supervised guidance facilitates hierarchical task decomposition, assisting the	natural language processing ,	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	agent to learn faster, achieve better driving performance and increases	speech processing and computer	Learning to drive faithfully in	learning to drive faithfully in	"Machine learning is the learning	In @cite, an end-to-end deep	Multi-task learning (MTL) @cite	In @cite, the authors proposed a
	transparency of the otherwise black-box end-to-end network. We run our	vision . In the field of sequential	highly stochastic urban settings	highly stochastic urban settings	of a mapping from situations to	neural network is used to learn	@cite aims to learn a mapping	multi-modal MTL method to
	experiments to validate the MT-LfD framework in CARLA, an open-source urban	decision making, @cite	remains an open problem. To	remains an open problem. this	actions so as to maximize a	to play Atari 2600 Atari games	from situations to actions so as	train a neural network to imitate
	driving simulator. We introduce multiple non-player agents in CARLA and induce	demonstrate MTL for 3D game	that end, we propose a Multi-	paper describes a technique for	scalar reward or reinforcement	directly from sensory experience.	to maximize a scalar reward or	the steering angle and driving
	temporal noise in them for realistic stochasticity.	playing, @cite and @cite	task Learning from	using Multi-Modal Multi-Task	signal. The learner is not told	The network is trained to imitate	reinforcement signal. The learner	speed of a car. In @cite, the
	(2):	demonstrate MTL in 3D maze	Demonstration (MT-LfD)	learning that considers multiple	which action to take, as in most	the steering angle and driving	is not told which action to take,	authors proposed an end-to-end
	An artificial agent is developed that learns to play a diverse range of classic Atari	navigation task whereas @cite	framework which uses	behavioral modalities as distinct	forms of machine learning, but	speed of human control of a car.	as in most forms of machine	trainable network for imitating
	2600 computer games directly from sensory experience, achieving a performance comparable to that of an expert human player; this work paves the	utilize the MTL framework for autonomous driving. Instead of	supervised auxiliary task prediction to guide the main task	modes of operation for an end- to-end autonomous deep neural	instead must discover which actions yield the highest reward	@cite proposed a multi-modal multi-task learning from	learning, but instead must discover which actions yield the	the expert demonstrator's driving commands. However,
	performance comparable to that of an expert human player; this work paves the way to building general-purpose learning algorithms that bridge the divide	employing future control outputs	of prediction to guide the main task	network utilizing the insertion of	by trying them. In the most	demonstration (MT-LfD)	highest reward by trying them. In	they do not provide a direct
	between perception and action.	as auxiliary tasks as shown by	commands. Our framework	modal information as secondary	interesting and challenging	framework to train an agent to	the most interesting and	supervision mechanism for the
	(3):	@cite . in this work we employ	involves an end-to-end trainable	input data for tasks with related	cases, actions may affect not	play Atari games.	challenging cases, actions may	agent to learn the driving
	Reinforcement learning is the learning of a mapping from situations to actions	action and visual abstractions to	network for imitating the expert	behaviors. using labeled data	only the immediate's reward, but		affect not only the immediate's	commands, which is not the case
	so as to maximize a scalar reward or reinforcement signal. The learner is not	guide the driving behavior.	demonstrator's driving	from hours of driving our fleet of	also the next situation, and	- Shows some sign of MDS,	reward, but also the next	in our case. Moreover, the
	told which action to take, as in most forms of machine learning, but instead		commands. The network	1 10th scale model cars, we	through that all subsequent	summarizing the 2nd (green)	situation, and through that all	authors do not propose a
	must discover which actions yield the highest reward by trying them. In the		intermediately predicts visual	trained multiple neural networks	rewards. These two	and 4th (purple) articles.	subsequent rewards @cite.	method to induce temporal
	most interesting and challenging cases, actions may affect not only the		affordances and action primitives	to imitate the steering angle and	characteristics—trial-and-error	<ul> <li>However, there is some mix up</li> </ul>	Comment:	noise in the network. In contrast,
	immediate's reward, but also the next situation, and through that all		through direct supervision which	driving speed of human control	search and delayed reward—are	about the networks for the two	- Summarized the 3rd article only	in our work, we propose a
	subsequent rewards. These two characteristics—trial-and-error search and delayed reward—are the two most important distinguishing features of		provide the aforementioned auxiliary supervised guidance.	of a car. we show that in each case, our	the two most important distinguishing features of	articles - The model then mentions the	- No indication of MDS	method that intermediately predicts the driving commands
	reinforcement learning.		We demonstrate that such joint	models trained with multi-modal	reinforcement learning." "In	1st article (yellow) but		to guide the agent to achieve
	(4):		learning and supervised	multi-task learning can match or	recent years different lines of	erroneously treated it as a non-		better driving performance and
	Several deep learning approaches have been applied to the autonomous driving		guidance facilitates hierarchical	outperform multiple networks	evidence have led to the idea	main article (red @cite) and		increase transparency.
	task, many employing end-to-end deep neural networks. Autonomous driving is		task decomposition, assisting the	trained on individual tasks, while	that motor actions and	that it is used to play Atari		Comment:
	complex, utilizing multiple behavioral modalities ranging from lane changing to		agent to learn faster, achieve	using a fraction of the	movements in both vertebrates	games instead of driving (2nd		- Shows some sign of MDS,
	turning and stopping. However, most existing approaches do not factor in the		better driving performance and	parameters and having more	and invertebrates are composed	red part)		contrasting the main article
	different behavioral modalities of the driving task into the training strategy. This		increases transparency of the	distinct mode of operation than	of elementary building blocks.			(yellow) with the other articles.
	paper describes a technique for using Multi-Modal Multi-Task Learning that		otherwise black-box end-to	a network trained without multi -	The entire motor repertoire can			- However, the only other article
	considers multiple behavioral modalities as distinct modes of operation for an		Comment:	modal multi- task learning on the	be spanned by applying a well-			explicitly mentioned is the 4th
	end-to-end autonomous deep neural network utilizing the insertion of modal information as secondary input data. Using labeled data from hours of driving		Copied the first tokens only     No indication of MDS	same data. Comment:	defined set of operations and transformations to these			one (purple).
	our fleet of 1 10th scale model cars, we trained multiple neural networks to		- NO INDICATION OF WIDS	- Started from the 1st sentence of	primitives and by combining			
	imitate the steering angle and driving speed of human control of a car. We show			the 1st article (yellow), but then	them in many different ways			
	that in each case, our models trained with MTL can match or outperform			moved to copy from the 3rd	according to well-defined			
	multiple networks trained on individual tasks, while using a fraction of the			sentence of the 4th article	syntactic rules. Motor and			
	parameters and having more distinct modes of operation than a network			(purple)	movement primitives and			
	trained without MTL on the same multi-modal data. These results should			- Words in red also came from	modules might exist at the			
	encourage Multi-Modal MTL-style training with the insertion of Modal			the 4th article but from a later	neural, dynamic and kinematic			
	Information for tasks with related behaviors.			part	levels with complicated mapping			
	(5): In recent years different lines of evidence have led to the idea that motor			- No indication of MDS	among the elementary building blocks subserving these different			
	actions and movements in both vertebrates and invertebrates are composed of				levels of representation. Hence.			
	elementary building blocks. The entire motor repertoire can be spanned by				while considerable progress has			
	applying a well-defined set of operations and transformations to these				been made in recent years in			
	primitives and by combining them in many different ways according to well-				unravelling the nature of these			
	defined syntactic rules. Motor and movement primitives and modules might				primitives, new experimental,			
	exist at the neural, dynamic and kinematic levels with complicated mapping				computational			
	among the elementary building blocks subserving these different levels of				Comment:			
	representation. Hence, while considerable progress has been made in recent				- Copied from the 3rd (purple)			
	years in unravelling the nature of these primitives, new experimental,				and 5 <sup>th</sup> (orange) articles only - No indication of MDS			
	computational and conceptual approaches are needed to further advance our understanding of motor compositionality.				- NO INGICATION OF MUS			
	anacistanang of motor compositionality.							
$\perp$								

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
845	(1):	A few extractive neural models	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	We present a neural framework for opinion summarization from online product	have been recently applied to	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	reviews which is knowledge-lean and only requires light supervision (e.g., in the	generic multi-document	0.0732 (prec)	0.0902 (prec)	0.0651 (prec)	0.1695 (prec)	0.1186 (prec)	0.1594 (prec)
	form of product domain labels and user-provided ratings). Our method	summarization. train a recursive	0.0889 (recall)	0.0815 (recall)	0.1037 (recall)	0.0741 (recall)	0.0519 (recall)	0.0815 (recall)
	combines two weakly supervised components to identify salient opinions and	neural network using a ranking	0.0803 (f-1)	0.0856 (f-1)	0.08 (f-1)	0.1031 (f-1)	0.0722 (f-1)	0.1078 (f-1)
	form extractive summaries from multiple reviews: an aspect extractor trained	objective to identify salient	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	under a multi-task objective, and a sentiment predictor based on multiple	sentences, while follow-up work	0.1394 (prec)	0.1707 (prec)	0.1157 (prec)	0.3333 (prec)	0.2167 (prec)	0.2429 (prec)
	instance learning. We introduce an opinion summarization dataset that includes	@cite employs a multi-task	0.1691 (recall)	0.1544 (recall)	0.1838 (recall)	0.1471 (recall)	0.0956 (recall)	0.125 (recall)
	a training set of product reviews from six diverse domains and human-	objective to improve sentence	0.1528 (f-1)	0.1622 (f-1)	0.142 (f-1)	0.2041 (f-1)	0.1327 (f-1)	0.165 (f-1)
	annotated development and test sets with gold standard aspect annotations,	extraction, an idea we adapted to	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	salience labels, and opinion summaries. Automatic evaluation shows significant	our task. propose a graph	We present a neural framework	we present a neural framework	We present a neural framework	In the context of multi-document	There is a large body of work on	In the context of summarization,
	improvements over baselines, and a large-scale study indicates that our opinion	convolution network to represent	for opinion summarization from	for opinion summarization from	for opinion summarization from	summarization, R2N2 @cite is a	multi-document summarization,	@cite employ a Graph
	summaries are preferred by human judges according to multiple criteria.	sentence relations and estimate	online product reviews which is	online product reviews which is	online product reviews which is	recurrent neural network (RNN)	including @cite @cite, @cite and	Convolutional Network (GCN) on
	(2):	sentence salience. Our	knowledge-lean and only	knowledge-lean and only	knowledge-lean and only	for extracting salience labels. The	@cite. However, most of these	the relation graphs, with
	The specification discloses a luggage carrier made up of a generally U-shaped	summarization method is tailored	requires light supervision (e.g., in	requires light supervision (e.g., in	requires light supervision (e.g., in	RNN is trained on sentence	works focus on extracting	sentence embeddings obtained
	frame. The frame has two spaced legs with a hook on the front which hooks	to the opinion extraction task, it	the form of product domain	the form of product domain	the form of product domain	relation graphs, and the salience	features from a single review. In	from Recurrent Neural Networks
	over the bumper of an automobile. Two braces are attached to the cross	identifies aspect-specific and	labels and user-provided ratings).	labels and user-provided	labels and user-provided ratings).	labels are extracted using a	this work, we focus on extractive	as input node features for
	member of the U-shaped member and the front portion of the braces is	salient units, while minimizing the	Our method combines two	ratings ).	Our method combines two	greedy heuristic to extract salient	summaries from multiple	salience estimation. However,
	received on fastening means welded to the under side of the car frame. The	redundancy of the final summary	weakly supervised components	our method combines two	weakly supervised components	sentences while avoiding	reviews: an aspect extractor	they do not consider sentence
	cross members provide a supporting surface for carrying articles, boats and the	with a greedy selection algorithm	to identify salient opinions and	weakly supervised components	to identify salient opinions and	redundancy @cite @cite. In this	trained under a multi-task	relation graphs and do not
	like. A platform may be supported on the frame.	@cite @cite . Redundancy is also	form extractive summaries from	to identify salient opinions and	form extractive summaries from	work, we use the RNN to extract	objective, and a sentiment	consider the representation
	(3):	addressed in who propose a	multiple reviews: an aspect	form extractive summaries from	multiple reviews: an aspect	salience labels from product	predictor trained based on	power of deep neural networks.
	We develop a Ranking framework upon Recursive Neural Networks (R2N2) to	graph-based framework for	extractor trained under a multi-	multiple reviews : an aspect	extractor trained under a multi-	reviews.	multiple instance learning.	In contrast, we use salience
	rank sentences for multi-document summarization. It formulates the sentence	abstractive summarization.	task objective, and a sentiment	extractor trained under a multi-	task objective, and a sentiment	Comment:	Comment:	labels and salience labels to
	ranking task as a hierarchical regression process, which simultaneously	introduce an encoder-decoder	predictor based on multiple	task objective, and a sentiment	predictor based on multiple	<ul> <li>Shows clear sign of MDS,</li> </ul>	<ul> <li>Shows some sign of MDS, with</li> </ul>	identify salient opinions and
	measures the salience of a sentence and its constituents (e.g., phrases) in the	neural method for extractive	instance learning. We introduce	predictor based on multiple	instance learning. We introduce	summarizing the 3rd (blue) and	the model summarizing what	form extractive summaries from
	parsing tree. This enables us to draw on word-level to sentence-level	opinion summarization. Their	an opinion summarization	instance learning.	an opinion summarization	4th (purple) articles, while	the main article does (yellow),	multiple reviews.
	supervisions derived from reference summaries. In addition, recursive neural	approach requires direct	dataset that includes a training	we introduce an opinion	dataset that includes a training	clearly contrasting the main	and trying to contrast that with	Comment:
	networks are used to automatically learn ranking features over the tree, with	supervision via gold-standard	set of product reviews from six	summarization dataset that	set of product reviews from six	article (yellow).	other literature (red)	- Shows some sign of MDS, the
	hand-crafted feature vectors of words as inputs. Hierarchical regressions are	extractive summaries for training,	diverse domains and human-	includes a training set of product	diverse domains and human-	<ul> <li>The RNN notation proved</li> </ul>	<ul> <li>However, the red part is</li> </ul>	model summarizing what the
	then conducted with learned features concatenating raw features. Ranking	in contrast to our weakly	annotated development and test	reviews from six diverse domains	annotated development and test	slightly confusing for the model	actually not factually based,	main article does (yellow), and
	scores of sentences and words are utilized to effectively select informative and	supervised formulation.	sets with gold standard aspect	and human-annotated	sets with gold standard aspect	though, the 2 <sup>nd</sup> article used	and the phrase "a large body of	trying to contrast that with
	non-redundant sentences to generate summaries. Experiments on the DUC		annotations, salience labels, and	development and test sets with	annotations, salience labels, and	"Recursive Neural Network",	work" is probably picked up	other literature (purple)
	2001, 2002 and 2004 multi-document summarization datasets show that R2N2		opinion summaries. Automatic	gold standard aspect	opinion summaries. Automatic	not "Recurrent Neural	from the summarization	- However, the red-and-purple
	outperforms state-of-the-art extractive summarization approaches.		evaluation shows significant	annotations, salience labels, and	evaluation shows significant	Network"	training.	part contains erroneous facts.
	(4):		improvements over baselines,	opinion summaries.	improvements over baselines,	<ul> <li>The 2<sup>nd</sup> article appears</li> </ul>		The model wrongly assumes
	We propose a neural multi-document summarization (MDS) system that		and a large-scale study indicates	a large-scale study indicates that	and a large-scale study indicates	completely unrelated to the		that the 4th article did not look
	incorporates sentence relation graphs. We employ a Graph Convolutional		that our opinion summaries are	our opinion summaries are	that our opinion summaries are	task, and the model probably		at sentence relation graphs
	Network (GCN) on the relation graphs, with sentence embeddings obtained		preferred by human judges	preferred by human judges	preferred by human judges	correctly omits it.		when the authors explicitly say
	from Recurrent Neural Networks as input node features. Through multiple layer-		according to multiple	according to multiple criteria.	according to multiple criteria.			they did.
	wise propagation, the GCN generates high-level hidden sentence features for		criteria.    The specification	Comment:	The specification discloses a			
	salience estimation. We then use a greedy heuristic to extract salient sentences		discloses a luggage carrier made	- Copied the first tokens only,	luggage carrier made up of a			
	while avoiding redundancy. In our experiments on DUC 2004, we consider three		up of a generally U-shaped	except that the phrase	generally U-shaped frame. The			
	types of sentence relation graphs and demonstrate the advantage of combining		frame. The frame has two spaced	"Automatic evaluation shows	frame has two spaced legs with a			
	sentence relations in graphs with the representation power of deep neural		legs with a hook on the front	significant improvements over	hook on the front which hooks			
	networks. Our model improves upon traditional graph-based extractive		which hooks over the bumper of	baselines, and" is skipped	over the bumper of an			
	approaches and the vanilla GRU sequence model with no graph, and it achieves		an automobile	before the red part	automobile. Two braces are			
	competitive results against other state-of-the-art multi-document		Comment:	- No indication of MDS	attached to the cross member of			
1	summarization systems.		- Copied the first tokens only		the U-shaped member and the			
			- No indication of MDS		front portion of the braces is			
					received on fastening means			
					welded to the under side of the			
1					car frame. The cross members			
					provide a supporting surface for			
					carrying articles, boats and the			
					like. A platform may			
					Comment:			
1					- Copied the first tokens only			
1 1					- No indication of MDS			
					I			

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
1165	(1):	As we noted above, most studies	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	In the age of social news, it is important to understand the types of reactions	that examine misinformation	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	that are evoked from news sources with various levels of credibility. In the	spread focus on individual events	0.0364 (prec)	0.0071 (prec)	0.028 (prec)	0.0156 (prec)	0.0182 (prec)	0.0129 (prec)
	present work we seek to better understand how users react to trusted and	such as natural disasters @cite,	0.0588 (recall)	0.0098 (recall)	0.0588 (recall)	0.0098 (recall)	0.0196 (recall)	0.0196 (recall)
	deceptive news sources across two popular, and very different, social media	political elections @cite, or crises	0.0449 (f-1)	0.0083 (f-1)	0.038 (f-1)	0.012 (f-1)	0.0189 (f-1)	0.0156 (f-1)
	platforms. To that end, (1) we develop a model to classify user reactions into	@cite and examine the response	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	one of nine types, such as answer, elaboration, and question, etc, and (2) we	to the event on social media. A	0.1084 (prec)	0.0993 (prec)	0.093 (prec)	0.1385 (prec)	0.0991 (prec)	0.0769 (prec)
	measure the speed and the type of reaction for trusted and deceptive news	recent study by Vosoughi al found	0.1748 (recall)	0.1359 (recall)	0.1942 (recall)	0.0874 (recall)	0.1068 (recall)	0.1165 (recall)
	sources for 10.8M Twitter posts and 6.2M Reddit comments. We show that	that news stories that were fact-	0.1338 (f-1)	0.1148 (f-1)	0.1258 (f-1)	0.1071 (f-1)	0.1028 (f-1)	0.0927 (f-1)
	there are significant differences in the speed and the type of reactions between	checked and found to be false	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	trusted and deceptive news sources on Twitter, but far smaller differences on	spread faster and to more people	In the age of social news, it is	social media have become an	In the age of social news, it is	In the context of crisis	In the past few years, there has	@cite use epidemiological
	Reddit.	than news items found to be true.	important to understand the	established feature of the	important to understand the	informatics, @cite and @cite use	been a growing interest in	models to characterize
	(2):	In contrast, our methodology	types of reactions that are	dynamic information space that	types of reactions that are	epidemiological models to	studying information diffusion on	information diffusion on social
	Social media have become an established feature of the dynamic information	considers immediate reactions to	evoked from news sources with	emerges during crisis events.	evoked from news sources with	characterize information	social media. For example, @cite	platforms like Twitter. They use
	space that emerges during crisis events. Both emergency responders and the	news of varying credibility, so we	various levels of credibility. In the	both emergency responders and	various levels of credibility. In the	cascades in Twitter resulting	used epidemiological models to	the SEIZ enhanced epidemic
	public use these platforms to search for, disseminate, challenge, and make	can determine whether certain	present work we seek to better	the public use these platforms to	present work we seek to better	from both news and rumors. In	characterize information	model that explicitly recognizes
	sense of information during crises. In these situations rumors also proliferate,	reactions or reactions to trusted	understand how users react to	search for, disseminate,	understand how users react to	@cite, the authors use the SEIZ	cascades in Twitter resulting	skeptics to characterize eight
	but just how fast such information can spread is an open question. We address	or deceptive news sources evoke	trusted and deceptive news	challenge, and make sense of	trusted and deceptive news	enhanced epidemic model that	from both news and rumors.	events across the world and
	this gap, modeling the speed of information transmission to compare	more or faster responses from social media users.	sources across two popular, and	information during crises. in	sources across two popular, and	explicitly recognizes skeptics to	@cite studied the spread of	spanning a range of event types.
	retransmission times across content and context features. We specifically	social media users.	very different, social media	these situations rumors also	very different, social media	characterize eight events across	misinformation on Twitter during	They do not explicitly recognize
	contrast rumor-affirming messages with rumor-correcting messages on Twitter		platforms. To that end, (1) we	proliferate, but just how fast	platforms. To that end, (1) we	the world and spanning a range	the 2017 French presidential	skeptics to characterize events in
	during a notable hostage crisis to reveal differences in transmission speed. Our work has important implications for the growing field of crisis informatics.		develop a model to classify user reactions into one of nine types,	such information can spread is an open question.	develop a model to classify user reactions into one of nine types,	of event types. They demonstrate that their approach	election. They collected a massive Twitter dataset of nearly	Twitter, and they do not consider the impact of rumors on the
	(3):		such as answer, elaboration, and	we address this gap by modeling	such as answer, elaboration, and	is accurate at capturing diffusion	17 million posts, posted between	spread of information diffusion
	Characterizing information diffusion on social platforms like Twitter enables us		question, etc, and (2) we	the speed of information	question, etc, and (2) we	in crisis informatics.	27 April and 7 May 2017	in Twitter. @cite study the use of
	to understand the properties of underlying media and model communication		measure the speed and the type	transmission to compare	measure the speed and the type	Comment:	(Election Day). They then set to	bots in the 2016 French
	patterns. As Twitter gains in popularity, it has also become a venue to broadcast		of reaction for trusted and	retransmission times across	of reaction for trusted and	- Shows no sign of MDS, with the	study the activities of the two	presidential election. They found
	rumors and misinformation. We use epidemiological models to characterize		deceptive information sources	content and context features.	misleading news sources for	model just summarizing the 3rd	groups independently, as well as	anomalous account usage
	information cascades in twitter resulting from both news and rumors.		for 10.8M Twitter posts and	we use epidemiological models	10.8M Twitter posts and 6.2M	article	their interplay. They provide a	patterns suggest the possible
	Specifically, we use the SEIZ enhanced epidemic model that explicitly recognizes		6.2M Reddit comments. We	to characterize information	Reddit comments. We show that	- Maybe this have something to	characterization of both the bots	existence of a black market for
	skeptics to characterize eight events across the world and spanning a range of		show that there are significant	cascades in twitter resulting from	there are significant differences	do with how the 1st article (i.e.	and the users who engaged with	reusable political disinformation
	event types. We demonstrate that our approach is accurate at capturing		differences in the speed, the	both news and rumors to	in the speed and a type of	the main one) is not clear what	them, and oppose it to those	bots, as well as their interplay
	diffusion in these events. Our approach can be fruitfully combined with other		type of reactions between	characterize eight events across	reactions between trusted and	it does?	who didn't.	with the users who engaged with
	strategies that use content modeling and graph theoretic features to detect		trusted and deceptive sites on	the world and spanning a range	deceptive News sources on		Comment:	them, and the users who had not
	(and possibly disrupt) rumors.		Twitter, but far smaller	of event types.	Twitter, but far smaller		- Shows some sign of MDS, with	engaged with them. The authors
	(4):		differences on Reddit.    Social	specifically, we contrast rumor-	differences on Reddit. Social		the model summarizing the 3 <sup>rd</sup>	of @cite found that the users
	Recent accounts from researchers, journalists, as well as federal investigators,		media have become an	affirming messages with rumor-	media has become an		(blue) and 4th (purple) articles	who did engage with the bots
	reached a unanimous conclusion: social media are systematically exploited to		established feature of the	correcting messages on Twitter	established feature of the		- The main article is probably not	are mostly foreigners, rather
	manipulate and alter public opinion. Some disinformation campaigns have been		dynamic information space that	during a notable hostage crisis to	dynamic information space that		mention because it does not include much detail? The 2 <sup>nd</sup>	than French users with diverse
	coordinated by means of bots, social media accounts controlled by computer scripts that try to disguise themselves as legitimate human users. In this study,		emerges during crisis events.  Both emergency responders and	reveal differences in transmission speed.	emerges during crisis events.  Both emergency responders and		article is also highly abstract	political views, and that the bots who engaged with the bots have
	we describe one such operation that occurred in the run up to the 2017 French		the public use these platforms to	our work has important	the public use these platforms to		article is also riigiliy abstract	diverse political views.
	presidential election. We collected a massive Twitter dataset of nearly 17 million		search for, disseminate,	implications for the growing field	search for, disseminate.			Comment:
	posts, posted between 27 April and 7 May 2017 (Election Day). We then set to		challenge, and make sense of	of crisis informatics.	challenge, and make sense of			- Shows some sign of MDS, with
	study the MacronLeaks disinformation campaign: By leveraging a mix of		Comment:	Comment:	information during crises. In			the model summarizing the 3rd
	machine learning and cognitive behavioral modeling techniques, we separated		- Copied the first tokens only	- Extracted and re-arranged	these situations rumors also			(blue) and 4th (purple) articles
	humans from bots, and then studied the activities of the two groups		- No indication of MDS	sentences from the 2nd (green)	proliferate, but just how fast			- Summarization is more clumsy
	independently, as well as their interplay. We provide a characterization of both			and 3rd (blue) articles only	such information can spread is			than Centrum
	the bots and the users who engaged with them, and oppose it to those users			- No indication of MDS	an open question. We address			- The 2017 French election is
	who didn't. Prior interests of disinformation adopters pinpoint to the reasons of				this gap, modeling the speed of			wrongly changed to 2016 (red)
1	scarce success of this campaign: the users who engaged with MacronLeaks are	1			information transmission to			despite 2016 is not mentioned
1	mostly foreigners with pre-existing interest in alt-right topics and alternative	1			compare retransmission times			anywhere else in the inputs
	news media, rather than French users with diverse political views. Concluding,	1			across content and context			
1	anomalous account usage patterns suggest the possible existence of a black	1			features. We specifically contrast			
	market for reusable political disinformation bots.	1			rumor-affirming messages with			
1		1			Comment: - Copied the first tokens only			
		1			Copied the first tokens only     No indication of MDS			
		1		l .	l .	l .		

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4263	(1):	Besides @cite , there are also	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Classical algorithms for query optimization presuppose the absence of	several works that automatically	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	inconsistencies or uncertainties in the database and exploit only valid semantic	derive and exploit soft (also called	0.0231 (prec)	0.0185 (prec)	0.0181 (prec)	0.023 (prec)	0.0357 (prec)	0.0209 (prec)
	knowledge provided, e.g., by integrity constraints. Data inconsistency or	dynamic) constraints for SQO	0.0755 (recall)	0.0189 (recall)	0.0755 (recall)	0.0377 (recall)	0.1132 (recall)	0.0755 (recall)
	uncertainty, however, is a widespread critical issue in ordinary databases: total	@cite @cite @cite , but this	0.0354 (f-1)	0.0187 (f-1)	0.0292 (f-1)	0.0286 (f-1)	0.0543 (f-1)	0.0328 (f-1)
	integrity is often, in fact, an unrealistic assumption and violations to integrity	semantic knowledge is used only	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	constraints may be introduced in several ways. In this report we present an	as long as it is true, and then	0.0747 (prec)	0.0909 (prec)	0.0676 (prec)	0.1136 (prec)	0.1006 (prec)	0.0677 (prec)
	approach for semantic query optimization that, differently from the traditional	either updated or discarded, so	0.2407 (recall)	0.0926 (recall)	0.2778 (recall)	0.1852 (recall)	0.3148 (recall)	0.2407 (recall)
	ones, relies on not necessarily valid semantic knowledge, e.g., provided by	that the information inferred from	0.114 (f-1)	0.0917 (f-1)	0.1087 (f-1)	0.1408 (f-1)	0.1525 (f-1)	0.1057 (f-1)
	violated or soft integrity constraints, or induced by applying data mining	the database by using it is correct.	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	techniques. Query optimization that leverages invalid semantic knowledge		Classical algorithms for query	an approach to learning query-	Classical algorithms for query	Semantic query optimization has	There is a large body of work on	In @cite, a rule deriver is used to
	cannot guarantee the semantic equivalence between the original user's query		optimization presuppose the	transformation rules based on	optimization suppose the	also been studied in the context	semantic query optimization	direct the search for rules that
	and its rewriting: thus a query optimized by our approach yields approximate		absence of inconsistencies or	analyzing the existing data in the	absence of inconsistencies or	of data mining @cite @cite. The	@cite @cite. In @cite, the	may be true only in the current
	answers that can be provided to the users whenever fast but possibly partial		uncertainties in the database	database is proposed.	uncertainties in the database	authors in @cite proposed a	authors present a framework for	state of the database <mark>. In</mark>
	responses are required. Also, we evaluate the impact of use of invalid semantic		and exploit only valid semantic	a framework and a closure	and exploit only valid semantic	framework for learning query-	learning query-transformation	contrast, our approach uses
	knowledge in the rewriting of a query by computing a measure of the quality of		knowledge provided, e.g., by	algorithm for learning rules from	knowledge provided, e.g., by	transformation rules based on	rules based on analyzing the	intermediate results from the
	the answer returned to the user, and we rely on the recent theory of Belief Logic		integrity constraints. Data	a given data distribution are	integrity constraints. Data	analyzing the existing data in the	existing data in the database.	optimization process to guide
	Programming to deal with the presence of possible correlation in the semantic		inconsistency or uncertainty,	described.	inconsistency or uncertainty,	database and learning rules that	The framework and a closure	the search for new rules that
	knowledge used in the rewriting.		however, is a widespread critical	the correctness, completeness,	however, is a widespread critical	may be true only in the current	algorithm for learning rules from	may not be true at all. In
	(2):		issue in ordinary databases: total	and complexity of the proposed	issue in ordinary databases: total	state of the database and can	a given data distribution are	addition, our system does not
1	The authors present an approach to acquiring knowledge from previously		integrity is often, in fact, an	algorithm are characterized and	integrity is often, in fact, an	modify the rule set to reflect	described. The approach is an	require intermediate results to
	processed queries. By using newly acquired knowledge together with given		unrealistic assumption and	a detailed example is provided to	unrealistic assumption and	changes in the data distribution.	extension of the EXODUS	select rules that are true only at the time of the search, and we
	semantic knowledge, it is possible to make the query processor and or optimizer more intelligent so that future queries can b processed more efficiently. The		violations to integrity constraints may be introduced in several	illustrate the framework. Comment:	violations to integrity constraints may be introduced in several	In @cite, the authors presented an approach to acquiring	conventional query optimizer generator @cite and an	do not require intermediate
	acquired knowledge is in the form of constraints. While some constraints are to		ways. In this report we present	Copied the 3rd article in full only	ways. In this report we present	knowledge from previously	extension of an extension of	rules to select the rules that
	be enforced for all database states, others are known to be valid for the current		an approach for semantic query	No indication of MDS	an approach for semantic query	processed queries and a method	conventional optimization in this	would or could be true. @cite
	state of the database. The former constraints are statistic integrity constraints,		optimization that, differently	- NO Indication of MD3	optimization that, differently	to maintain the knowledge base	context. The authors in @cite	proposed a method to compute
	while the latter are called dynamic integrity constraints. Some situations in		from the traditional ones, relies		from the traditional ones, relies	is presented.	present an approach to acquiring	the restriction closure, i.e. all
	which certain dynamic semantic constraints can be automatically extracted are		on not necessarily valid semantic		on not necessarily valid semantic	Comment:	knowledge from previously	deductible restrictions, from a
	identified. This automatic tool for knowledge acquisition can also be used as an		knowledge, e.G., provided by		knowledge, e.e., provided by	- Shows some sign of MDS, with	processed queries. By using	given set of restrictions, join
	interactive tool for identifying potential static integrity constraints. The concept		violated or soft integrity		violated or soft integrity	the model summarizing the 3rd	newly acquired knowledge	predicates (as given in a query),
	of minimal knowledge base is introduced, and a method to maintain the		constraints, or induced by		constraints, or induced by	(blue) and 2nd (green) articles	together with given semantic	and constraints. The algorithm to
	knowledge base is presented. An algorithm to compute the restriction		applying data mining techniques.		applying data mining techniques.	- Interestingly, the yellow part is	knowledge, it is possible to make	compute the closure is given,
	(selection) closure, i.e. all deductible restrictions, from a given set of		Query optimization that		The correctness, completeness,	adopted from the 1st article,	the query processor and or	and a method to maintain the
	restrictions, join predicates (as given in a query), and constraints is given. >		leverages invalid semantic		and complexity of the proposed	but represents an inferred	optimizer more intelligent so	knowledge base is presented.
	(3):		knowledge cannot guarantee the		algorithm are characterized and	piece of information, i.e. that	that future queries can b	However, the method to
	An approach to learning query-transformation rules based on analyzing the		semantic equivalence between		a detailed example is provided to	datamining is one the	processed more efficiently. The	maintain knowledge base is not
	existing data in the database is proposed. A framework and a closure algorithm		the original user's query and its		illustrate the framework. >The	traditional (i.e. previously seen)	concept of minimal knowledge	presented in this paper. The
	for learning rules from a given data distribution are described. The correctness,		rewriting: thus a query optimized		use of inference rules to support	methods	base is introduced, and a	authors present an approach to
	completeness, and complexity of the proposed algorithm are characterized and		by our approach yields		intelligent data processing is an		method to maintain the	acquiring knowledge from
	a detailed example is provided to illustrate the framework. >		approximate answers that can be		increasingly important tool in		knowledge base is presented. An	previously processed queries,
	(4):		provided to the users whenever		many areas of computer science.		algorithm to compute the	which is based on the notion of
	The use of inference rules to support intelligent data processing is an		fast but possibly partial		In database systems, rules are		restriction (selection) closure, i.e.	minimal knowledge base.
	increasingly important tool in many areas of computer science. In database		responses are required. Also, we evaluate the impact of use of		used in semantic query optimization as a method for		all deductible restrictions, from a	However, this approach is not based on the minimal knowledge
	systems, rules are used in semantic query optimization as a method for reducing		invalid semantic knowledge in				given set of restrictions, join	
	query processing costs. The savings is dependent on the ability of experts to supply a set of useful rules and the ability of the optimizer to quickly find the		the rewriting of a query by		reducing query processing costs. The savings is dependent on the		predicates (as given in a query), and constraints is given.	base, and it is not possible to maintain any knowledge base.
	appropriate transformations generated by these rules. Unfortunately, the most		computing a measure of the		ability of experts to supply a set		Comment:	Comment:
	useful rules are not always those that would or could be specified by an expert.		quality of the answer returned		of useful rules and the ability of		- Shows clear sign of MDS,	- Shows some sign of MDS, with
	This paper describes the architecture of a system having two interrelated		Comment:		the optimizer to quickly find the		successfully summarizing all of	the model summarizing the 2 <sup>nd</sup>
	components: a combined conventional semantic query optimizer, and an		- Copied the first tokens only		appropriate transformations		the 2 <sup>nd</sup> (green), 3 <sup>rd</sup> (blue) and	(green) and 4th (purple) articles
	automatic rule deriver. Our automatic rule derivation method uses intermediate		- No indication of MDS		generated by these rules.		4th (red) articles while	- The 1st red part shows an
	results from the optimization process to direct the search for learning new rules.				Unfortunately, the most useful		mentioning the context	attempt to contrast with the
	Unlike a system employing only user-specified rules, a system with an automatic				rules are not always those that		(yellow).	main article but wrongly mixed
1	capability can derive rules that may be true only in the current state of the				would or could be specified by		- The 1st red part ("large body of	that up with facts from the 4th
	database and can modify the rule set to reflect changes in the database and its				an expert. This paper describes		work") is also used in some	one
	usage pattern. This system has been implemented as an extension of the				the architecture of a system		other summaries, possibly a	- The 2 <sup>nd</sup> and 3 <sup>rd</sup> red parts are
1	EXODUS conventional query optimizer generator. We describe the				having two interrelated		writing style picked up by the	pure hallucinations.
1	implementation, and show how semantic query optimization is an extension of				components		model.	
	conventional optimization in this context.				Comment:		- The 2 <sup>nd</sup> red part contains a typo	
					- Copied the 1st article in full,		("can b") that was directly	
					then the last sentence of the		copied by the model.	
					3 <sup>rd</sup> (blue) and then the rest of the 4 <sup>th</sup> (purple)			
1					the 4 <sup>th</sup> (purple)  - No indication of MDS			
					NO MAICEMON OF WIDS			
Ь					l .			

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4371	(1):	Similar to XRANK @cite is the	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	In contrast to XML query languages as e.g. XPath which require knowledge on	stack-based approach presented	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	the query language as well as on the document structure, keyword search is	in @cite . In contrast to the	0.0244 (prec)	0.0 (prec)	0.0061 (prec)	0.0429 (prec)	0.125 (prec)	0.04 (prec)
	open to anybody. As the size of XML sources grows rapidly, the need for efficient	previous stack-based appraoches,	0.069 (recall)	0.0 (recall)	0.0172 (recall)	0.0517 (recall)	0.0517 (recall)	0.0862 (recall)
	search indices on XML data that support keyword search increases. In this	the authors do not used the	0.036 (f-1)	0.0 (f-1)	0.009 (f-1)	0.0469 (f-1)	0.0732 (f-1)	0.0546 (f-1)
	paper, we present an approach of XML keyword search which is based on the	DeweyID to identify a node and to	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	DAG of the XML data, where repeated substructures are considered only once,	calculate the ancestor-descendant	0.0788 (prec)	0.1282 (prec)	0.0793 (prec)	0.1831 (prec)	0.4 (prec)	0.1349 (prec)
	and therefore, have to be searched only once. As our performance evaluation	or even parent-child relationships,	0.2203 (recall)	0.1695 (recall)	0.2203 (recall)	0.2203 (recall)	0.1695 (recall)	0.2881 (recall)
	shows, this DAG-based extension of the set intersection search algorithm[1], [2],	but they propose to use a	0.1161 (f-1)	0.146 (f-1)	0.1166 (f-1)	0.2 (f-1)	0.2381 (f-1)	0.1838 (f-1)
	can lead to search times that are on large documents more than twice as fast as	combination of preorder position,	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	the search times of the XML-based approach. Additionally, we utilize a smaller	postorder position, and depth of	In contrast to XML query	in this paper, we consider the	"We consider the problem of	XRANK @cite is an extension of	XRANK @cite is an extension of	XRANK @cite is a hyperlink
	index, i.e., we consume less main memory to compute the results.	the node.	languages as e.g. XPath which	problem of efficiently producing	efficiently producing ranked	the set intersection search	the set intersection search	based HTML search engine. It is
	(2):		require knowledge on the query	ranked results for keyword	results for keyword search	algorithm @cite, which is based	algorithm @cite. XRANK can be	designed to perform keyword
	Recent works have shown the benefits of keyword proximity search in querying		language as well as on the	search queries over hyperlinked	queries over hyperlinked Xie	on the DAG of the XML data.	used to query a mix of HTML and	search over hyperlinked XML
	XML documents in addition to text documents. For example, given query		document structure, keyword	XML documents.	documents, as opposed to	XRANK is a hyperlink based	XML documents.	documents. However, it does not
	keywords over Shakespeare's plays in XML, the user might be interested in		search is open to anybody. As	we present an approach of XML	(conceptually) flat HTML	HTML search engine that can be	Comment:	consider the nested structure of
	knowing how the keywords cooccur. In this paper, we focus on XML trees and		the size of XML sources grows	keyword search which is based	documents, introduces many	used to query a mix of hyperlink	- Shows very weak signs of MDS,	the XML documents. XRANK can
	define XML keyword, proximity queries to return the (possibly heterogeneous)		rapidly, the need for efficient	on the set intersection search	new challenges," the authors	based XML documents. X-RANK	with an attempt to reference	be used to query a mix of HTML
	set of minimum connecting trees (MCTs) of the matches to the individual		search indices on XML data that	algorithm, where repeated	write. "First, Xie XRANK does not	is designed to handle the	multiple @cite.	and XML documents, but it does
	keywords in the query. We consider efficiently executing keyword proximity		support keyword search	substructures are considered	always return entire documents,	problem of efficiently producing	- Even worse than tuned LED	not require a hyperlink search
	queries on labeled trees (XML) in various settings: 1) when the XML database		increases. In this paper, we	only once, and therefore, have to	but can return deeply nesting	ranked results for keyword	- However, none of the	engine, and it does not need a
	has been preprocessed and 2) when no indices are available on the XML		present an approach of XML	be searched only once.	markup that contains the desired	search queries over hyperlinked	sentences generated are	hyperlink, and it is not required
	database. We perform a detailed experimental evaluation to study the benefits		keyword search which is based	we perform a detailed	keywords. Second, the nesting	XML documents, as opposed to	coherent, showing difficulties	to search for hyperlinks in a
	of our approach and show that our algorithms considerably outperform prior		on the DAG of the XML data.	experimental evaluation to study	structure of markup implies that	(conceptually) flat HTML	of the model when dealing	hyperlink document. Moreover,
	algorithms and other applicable approaches.		where repeated substructures	the benefits of our approach and	the notion of ranking is no longer	documents.	with articles filled with	XRANK is designed to handle the
	(3):		are considered only once, and	show that our algorithms	at the granularity of a document,	Comment:	acronyms (XML, HTML) which it	nested structures of the XML
	We consider the problem of efficiently producing ranked results for keyword		therefore, have to be searched	considerably outperform prior	but at the granular of an markup.	- Shows very weak signs of MDS,	does not understand	document, In contrast, our
	search queries over hyperlinked XML documents. Evaluating keyword search		only once. As our performance	algorithms and other applicable	Finally, the notion of keyword	with an attempt to reference		approach is based on a DAG-
	queries over hierarchical XML documents, as opposed to (conceptually) flat		evaluation shows, this DAG-	approaches.	proximity is more complex in the	multiple @cite.		based extension of the set
	HTML documents, introduces many new challenges. First, XML keyword search		based extension of the set	Comment:	hierarchical markup data model.	- However, none of the		intersection search algorithm,
	queries do not always return entire documents, but can return deeply nested		intersection search algorithm[1],	- Very weak signs of MDS, with	In this paper, we present the	sentences generated are		which can lead to search times
	XML elements that contain the desired keywords. Second, the nested structure		[2], can lead to search times that	all 3 articles covered.	XRANK system that is designed	coherent, showing difficulties		that are more than twice as fast
	of XML implies that the notion of ranking is no longer at the granularity of a		are on large documents more	- However, they are all presented	to handle these novel features of	of the model when dealing		as the search time of the XML-
	document, but at the granularity of an XML element. Finally, the notion of		than twice as fast as the search	as a single document.	Xie keyword search. Our	with articles filled with		based approach.
	keyword proximity is more complex in the hierarchical XML data model. In this		times of the XML-based	- The model also performs like	experimental results show that	acronyms (XML, HTML) which it		Comment:
	paper, we present the XRANK system that is designed to handle these novel		approach. Additionally, we utilize	an extractive model	XRANK offers both space and	does not understand		- Shows signs of MDS, with the
	features of XML keyword search. Our experimental results show that XRANK		a smaller index, i.e., we consume		performance benefits when			yellow part contrasting the
	offers both space and performance benefits when compared with existing		less main memory to compute		compared with existing			approach of the main article
	approaches. An interesting feature of XRANK is that it naturally generalizes a		the results.               Recent works		approaches. An interesting			with that of the 3rd (blue)
	hyperlink based HTML search engine such as Google. XRANK can thus be used		have shown the benefits of		feature of XRANK is that it			- However, the red parts still
	to query a mix of HTML and XML documents.		keyword proximity search in		naturally generalizes a hyperlink			show hallucination
1			querying XML documents in	I	based HTML search engine such			The first step LED model in the
			addition to text documents.		as Google. XRANK can thus be			two-step model probably
1			Comment:	I	used to guery a mix of HTML and			helped simplify the meaning of
1			- Copied the first tokens only	I	markup markup."			the individual passages?
			- No indication of MDS		Comment:			posseges.
1				I	- Copied the 3rd article in full			
1			1	I	- No indication of MDS			
1			1	I				
1			1	I	I			
1	I .	1	I	I	1	I	1	1

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4717	(1):	The methods for constructing	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	One of the most influential recent results in network analysis is that many	graphs with a given degree	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	natural networks exhibit a power-law or log-normal degree distribution. This	distribution are primarily either	0.0414 (prec)	0.1398 (prec)	0.0714 (prec)	0.1607 (prec)	0.129 (prec)	0.097 (prec)
	has inspired numerous generative models that match this property. However,	reductions to perfect matchings	0.038 (recall)	0.0707 (recall)	0.0815 (recall)	0.0489 (recall)	0.0435 (recall)	0.0707 (recall)
	more recent work has shown that while these generative models do have the	or sequential sampling methods.	0.0397 (f-1)	0.0939 (f-1)	0.0761 (f-1)	0.075 (f-1)	0.065 (f-1)	0.0818 (f-1)
	right degree distribution, they are not good models for real life networks due to	There are two popular perfect	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	their differences on other important metrics like conductance. We believe this	matching methods. The first is the	0.1529 (prec)	0.2128 (prec)	0.1422 (prec)	0.3158 (prec)	0.2698 (prec)	0.2 (prec)
	is, in part, because many of these real-world networks have very different joint	@cite @cite : @math mini-	0.1405 (recall)	0.1081 (recall)	0.1622 (recall)	0.0973 (recall)	0.0919 (recall)	0.1459 (recall)
	degree distributions, i.e. the probability that a randomly selected edge will be	vertices are created for each	0.1465 (f-1)	0.1434 (f-1)	0.1515 (f-1)	0.1488 (f-1)	0.1371 (f-1)	0.1688 (f-1)
	between nodes of degree k and l. Assortativity is a sufficient statistic of the joint	degree @math vertex, and all the	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	degree distribution, and it has been previously noted that social networks tend	mini-vertices are connected. Any	One of the most influential	we propose a random graph	In a paper published in the	In @cite, the authors show that	The problem of generating	In @cite, the authors presented
	to be assortative, while biological and technological networks tend to be	perfect matching in the	recent results in network analysis	model which is a special case of	journal Random Struct, John	the joint degree distribution of	graphs with the same degree	a random graph model which is
	disassortative. We suggest understanding the relationship between network	configuration graph corresponds	is that many natural networks	sparse random graphs with given	Wiley & Sons, Inc. Random	graphs with a given degree	distribution was first studied by	based on the log-log growth rate
	structure and the joint degree distribution of graphs is an interesting avenue of	to a graph with the correct degree	exhibit a power-law or log-	degree sequences.	Struct. Alg., 14: 293-308, 1999,	distribution is asymptotic to the	Jerrum and Sinclair @cite. They	of the graph. However, the
	further research. An important tool for such studies are algorithms that can	distribution by merging all of the	normal degree distribution. This	this model involves only a small	the authors explain the	number of labelled @math	showed that the asymptotic	authors do not consider the
	generate random instances of graphs with the same joint degree distribution.	identified mini-vertices. This	has inspired numerous	number of parameters, called	relationship between network	-regular graphs on @math	distribution of the number of	distribution of the connected
	This is the main topic of this paper and we study the problem from both a	allows multiple edges and self-	generative models that match	logsize and log-log growth rate,	structure and the joint degree	vertices. The authors also show	short cycles in graphs with a	components of the graph, which
	theoretical and practical perspective. We provide an algorithm for constructing	loops, which are often	this property. However, more	which capture some universal	distribution of graphs. They	that their model is rapidly	given degree sequence, and gave	is not the case in the case in
1	simple graphs from a given joint degree distribution, and a Monte Carlo Markov	undesirable. See Figure . The	recent work has shown that	characteristics of massive graphs.	propose a random graph model	mixing, if the score sequence is	analogous formulae for	@cite. In @cite, the authors
1	Chain method for sampling them. We also show that the state space of simple	second approach, the , prevents	while these generative models	we also show that the state	which is a special case of sparse	near-regular. However, they do	hypergraphs. Later, @cite	presented two problems:
	graphs with a fixed degree distribution is connected via end point switches. We	multi-edges and self-loops by	do have the right degree	space of simple graphs with a	random graphs with given	not provide a proof of rapid	showed that their Markov chain	randomly generating bipartite
1	empirically evaluate the mixing time of this Markov Chain by using experiments	creating a gadget for each vertex.	distribution, they are not good models for real life networks due	fixed degree distribution is	degree sequences. This model	mixing.	on regular graphs is rapidly	graphs with a given degree
	based on the autocorrelation of each edge. These experiments show that our	If @math has degree @math ,		connected via end point	involves only a small number of	Comment:	mixing, if the score sequence is	sequence and randomly
	Markov Chain mixes quickly on real graphs, allowing for utilization of our	then it is replaced with a	to their differences on other	switches. as a consequence of	parameters, called logsize and	- Shows very weak signs of MDS,	near-regular. Comment:	generating tournaments with a
	techniques in practice.	complete bipartite graph @math	important metrics like	the method we determine the	log-log growth rate. These	with an attempt to provide		given score sequence. In the first
	(2):	with @math and @math . Exactly	conductance. We believe this is,	asymptotic distribution of the	parameters capture some universal characteristics of	some facts on the 3 <sup>rd</sup> (blue) and 4 <sup>th</sup> articles.	- Shows very weak signs of MDS,	problem, they can prove that
	We propose a random graph model which is a special case of sparse random	one node in each @math is	in part, because many of these	number of short cycles in graphs		4 <sup>th</sup> articles.	with an attempt to provide	their chains are rapidly mixing in
	graphs with given degree sequences. This model involves only a small number of parameters, called logsize and log-log growth rate. These parameters capture	connected to each other @math , representing edge @math @cite .	real-world networks have very different joint degree	with a given degree sequence, and give analogous formulae for	massive graphs. Furthermore, from these parameters, various		some facts on the 3 <sup>rd</sup> (blue) and 4 <sup>th</sup> articles.	general, but in the near-regular case, they cannot prove the
	some universal characteristics of massive graphs. Furthermore, from these	Any perfect matching in this	distributions, i.e. the probability	hypergraphs.	properties of the graph can be		4 drucies.	Markov chain in general, since
	parameters, various properties of the graph can be derived. For example, for	model corresponds exactly to a	that a randomly selected edge	Comment:	derived. For example, for certain			the score sequence of the two
	certain ranges of the parameters, we will compute the expected distribution of	simple graph by using the edges in	will be between nodes of degree	- Very weak signs of MDS, with	ranges of the parameters, we will			problems is near-normal. The
	the sizes of the connected components which almost surely occur with high	the matching that correspond	k and l. Assortativity is a	the 1st (yellow), 2nd (green) and	compute the expected			proof techniques we use for the
	probability. We will illustrate the consistency of our model with the behavior of	with edges connecting any	sufficient statistic of the joint	3rd (blue) articles covered but	distribution of the sizes of the			two problems are similar to the
	some massive graphs derived from data in telecommunications. We will also	@math to any @math . We use a	degree distribution, and it has	not the 4th.	connected components which			one used by Jerrum and Sinclair
	discuss the threshold function, the giant component, and the evolution of	natural extension of the first	been previously noted that social	- However, they are all presented	almost surely occur with high			@cite. However, they are
	random graphs in this model.	configuration model to the joint	networks tend to have	as a single document.	probability. We will illustrate the			different.
	(3):	degree distribution problem.	assortative, while biological and	- The model also performs like	consistency of our model with			Comment:
	Let $\Delta$ and n be natural numbers such that $\Delta n = 2m$ is even and $\Delta \leq (2 \log n) \cdot 12$ .		technological networks tend to	an extractive model	the behavior of some massive			- Shows signs of MDS, covering
	<ol> <li>Then as n →, the number of labelled Δ-regular graphs on n vertices is</li> </ol>		be disassortative. We suggest		graphs derived from data in			elements of the 2 <sup>nd</sup> (green), 3 <sup>rd</sup>
	asymptotic to $e - \lambda - \lambda 2 (2 \text{ m})! \text{ m}! 2 \text{ m} (\Delta!) \text{ m}$ where $\lambda = (\Delta - 1) 2$ . As a		understanding the relationship		telecommunications. We will			(blue) and 4th (purple) articles
	consequence of the method we determine the asymptotic distribution of the		between network structure and		also discuss the threshold			- However, the first red part is
	number of short cycles in graphs with a given degree sequence, and give		the joint degree distributions of		function, the giant component,			still factually incorrect while
	analogous formulae for hypergraphs.		graphs is an interesting avenue		and the evolution of random			the 2nd red part is quite unclear
	(4):		of further research. An		graphs in this model. For the first			- The first step LED model in the
	We consider two problems: randomly generating labeled bipartite graphs with a		important tool for such studies		problem, we cannot prove that			two-step model probably
	given degree sequence and randomly generating labeled tournaments with a		are algorithms that		our chain is rapidly mixing in			helped simplify the meaning of
	given score sequence. We analyze simple Markov chains for both problems. For		Comment:		general, but in the near-regular			the individual passages?
	the first problem, we cannot prove that our chain is rapidly mixing in general,		- Copied the first tokens only		case, i.e., when all the degrees			
	but in the near-regular case, i.e., when all the degrees are almost equal, we give		- No indication of MDS		are almost equal, we give a proof			
	a proof of rapid mixing. Our methods also apply to the corresponding problem				of rapid mixing. Our methods			
1	for general (nonbipartite) regular graphs, which was studied earlier by several	1			also apply to the corresponding	l		
	researchers. One significant difference in our approach is that our chain has one				problem for general			
1	state for every graph (or bipartite graph) with the given degree sequence; in	1			(nonbipartite) regular graphs,	l		
1	particular, there are no auxiliary states as in the chain used by Jerrum and	1			which was studied	l		
1	Sinclair. For the problem of generating tournaments, we are able to prove that	1			Comment:	I		
1	our Markov chain on tournaments is rapidly mixing, if the score sequence is	1			- Very weak signs of MDS, with	I		
1	near-regular. The proof techniques we use for the two problems are similar.	1			the 1st (yellow), 2nd (green) and	I		
1	©1999 John Wiley & Sons, Inc. Random Struct. Alg., 14: 293–308, 1999	1			4th (purple) articles covered but	I		
1		1			not the 3rd.	I		
1		1			- However, they are all presented	I		
1		1			as a single document.  - The model also performs like	I		
1		1			an extractive model	I		
1		1			an extractive moder	I		
1	I	1	1	l	I	I	1	1

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
4858	(1):	Concerning set packings the	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	It is shown that one can count @math -edge paths in an @math -vertex graph	situation is analogous, albeit the	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	and @math -set @math -packings on an @math -element universe,	research has been somewhat less	0.1079 (prec)	0.0612 (prec)	0.0791 (prec)	0.125 (prec)	0.082 (prec)	0.0745 (prec)
	respectively, in time @math and @math , up to a factor polynomial in @math ,	extensive. Deciding whether a	0.1442 (recall)	0.0577 (recall)	0.1346 (recall)	0.0769 (recall)	0.0962 (recall)	0.0673 (recall)
	@math , and @math ; in polynomial space, the bounds hold if multiplied by	given family of @math subsets of	0.1235 (f-1)	0.0594 (f-1)	0.0996 (f-1)	0.0952 (f-1)	0.0885 (f-1)	0.0707 (f-1)
	@math or @math , respectively. These are implications of a more general	an @math -element universe	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	result: given two set families on an @math -element universe, one can count	contains a @math -packing is	0.1714 (prec)	0.2222 (prec)	0.1348 (prec)	0.2923 (prec)	0.1951 (prec)	0.1895 (prec)
	the disjoint pairs of sets in the Cartesian product of the two families with	known to be W[1]-hard @cite ,	0.2286 (recall)	0.2095 (recall)	0.2286 (recall)	0.181 (recall)	0.2286 (recall)	0.1714 (recall)
	@math basic operations, where @math is the number of members in the two	and thus it is unlikely that the	0.1959 (f-1)	0.2157 (f-1)	0.1696 (f-1)	0.2235 (f-1)	0.2105 (f-1)	0.18 (f-1)
	families and their subsets.	problem is fixed parameter	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	(2):	tractable, that is, solvable in time	It is shown that one can count	given a set @math with @math	It is shown that one can count	For the Steiner tree problem	In @cite, Bjorklund and Husfeldt	In @cite, Dreyfus-Wagner
	We present a fast algorithm for the subset convolution problem: given functions	@math for some function @math	@math -edge paths in a @math	elements and a family @math of	@math -edge paths in an @math	@cite, one can count the disjoint	showed that one can partition a	algorithm @cite was presented,
	f and g defined on the lattice of subsets of ann-element set n, compute their	and constant @math . If @math is	-vertex graph and @math -set	subsets, we show how to	-vertex graph and @math -set	pairs of sets in the Cartesian	set @math into @math subsets	which is based on an @math
	subset convolution f*g, defined for $S\subseteq N$ by [ $(f * g)(S) = [T \subseteq S]$ f(T) g(S	fairly large, say exponential in	@math -packings on an @math	partition @math into @math	@math -packings on an @math-	product of the two families with	in @math time, where @math is	time bound of the classical
	T),,]where addition and multiplication is carried out in an arbitrary ring. Via	@math , the fastest known	-element universe, respectively,	such subsets in @math time.	element universe, respectively, in	@math basic operations, where	the number of members in the	DREW algorithm. The algorithm
	Mobius transform and inversion, our algorithm evaluates the subset convolution	algorithms actually count the	in time @math and @math, up	we also consider variations of	time @math and @math, up to a	@math is the number of	two families and their subsets.	was extended to the Steiner Tree
	in O(n2 2n) additions and multiplications, substanti y improving upon the	packings by employing the	to a factor polynomial in @math,	this problem where the subsets	factor polyn Chocobo in @math,	members in the two families and	They also showed how to	problem by @cite and presented
	straightforward O(3n) algorithm. Specifically, if the input functions have	inclusionexclusion machinery	@math, and @math ; in	may overlap or are weighted,	@math, and @math; in polyn	their subsets @math. For the	partition the set into @math	an O(3n - @math ) time bound
	aninteger range [-M,-M+1,,M], their subset convolution over the ordinary	@cite @cite and run in time	polynomial space, the bounds	and we solve the decision,	Chocobo space, the bounds hold	subset convolution problem, one	such subsets in time @math,	for the subgraphs with bounded
	sumproduct ring can be computed in O(2n log M) time; the notation O	@math . This bound holds also for	hold if multiplied by @math or	counting, summation, and	if multiplied by @math or	can count @math -edge paths in	where @math and @math are	integer weights. In the case
	suppresses polylogarithmic factors. Furthermore, using a standard embedding	the presented algorithm (cf.	@math, respectively. These are	optimization versions of these	@math, respectively. These are	an @math -vertex graph and	the subsets of @math. They also	where @math is not available,
	technique we can compute the subset convolution over the maxsum or min	Theorem ).	implications of a more general	problems.	implications of a more general	@math -set @math -packings in	gave a family of polynomial	@cite showed how to partition
	sum semiring in O(2n M) time. To demonstrate the applicability of fast subset		result: given two set families on	our algorithms are based on the	result: given two set families on	time @math.	space approximation algorithms	@math into @math such subsets
	convolution, wepresent the first O(2k n2 + n m) algorithm for the Steiner tree		an @x-element universe, one	principle of inclusion-exclusion	an @ math -element universe,	Comment:	that find a number between	in polynomial time and @math
	problem in graphs with n vertices, k terminals, and m edges with bounded		can count the disjoint pairs of	and the zeta transform.	one can count the disjoint pairs	<ul> <li>No indication of MDS behavior,</li> </ul>	@math and the @math in	in time @cite. In this case,
	integer weights, improving upon the O(3kn + 2k n2 + n m) time bound of the		sets in the Cartesian product of	in effect we get exact algorithms	of sets in the Cartesian product	basically just mixing in some	@math. @cite gave a fast	@math and @math are
	classical Dreyfus-Wagner algorithm. We also discuss extensions to recent O(2n)-		the two families with @math	in time @math for several well -	of the two families with @math	words of the 2 <sup>nd</sup> article (green)	algorithm for the subset	polynomial in time, and @cite
	time algorithms for covering and partitioning problems (Bjorklund and Husfeldt,		basic operations, where @math	studied partition problems	basic operations, where @math	with the 1st (yellow)	convolution problem in graphs	gives an approximation
	FOCS 2006; Koivisto, FOCS 2006).		is the number of members in the	including domatic number,	is the number of members in the	- Shows once again how the	with @math vertices, @math	algorithm for chromatic number
	(3):		two families and their	chromatic number, maximum	two families and their subsets.	model fails to understand	terminals, and @math edges	and domatic number.
	Given a set @math with @math elements and a family @math of subsets, we		subsets.     We present a fast	@math -cut, bin packing, list	We present a fast algorithm for	words used in unusual settings	with bounded integer weights,	Comment:
	show how to partition @math into @math such subsets in @math time. We		algorithm for the subset	coloring, and the chromatic	the subset convolution	(e.g. mathematics)	improving upon the classical	- Best performing of all 3 tuned
	also consider variations of this problem where the subsets may overlap or are		convolution problem:given	polynomial.	problem:given functions f and g		Dreyfus-Wagner algorithm. They	models, showing strong signs
	weighted, and we solve the decision, counting, summation, and optimization		functions f and g defined on the	Comment:	defined on the lattice of subsets		also discussed extensions to	of MDS with the 2nd (green)
	versions of these problems. Our algorithms are based on the principle of		lattice of subsets of ann-element	- Copied all of the 3 <sup>rd</sup> article	of ann-element set n, compute		recent @math -time algorithms	and 3rd (blue) articles reflected.
	inclusion-exclusion and the zeta transform. In effect we get exact algorithms in		set n, compute their subset	except for last 2 sentences	their subset convolution f*g,		for covering and partitioning	- The red part appears to contain
	@math time for several well-studied partition problems including domatic		convolution f*g, defined for S⊆	- No indication of MDS	defined for $S\subseteq N$ by [ $(f * g)(S) =$		problems.	some factual errors, with the
	number, chromatic number, maximum @math -cut, bin packing, list coloring,		Comment:		[T⊆S] f(T) g(ST),]where		Comment:	original "O(3kn + 2k n2 + n m)"
	and the chromatic polynomial. We also have applications to Bayesian learning		- Copied the first tokens only		addition and multiplication is		- Some weak signs of MDS, with	changed to "O(3n - @math )"
	with decision graphs and to model-based data clustering. If only polynomial	1	- No indication of MDS		carried out in an arbitrary ring.		the 2 <sup>nd</sup> (green) and 3 <sup>rd</sup> (blue)	- The first step LED model in the
	space is available, our algorithms run in time @math if membership in @math	1	- Random word change noted		Via Mobius transform and		articles referenced and mixed	two-step model probably
1	can be decided in polynomial time. We solve chromatic number in @math time	1	(red)	I	inversion, our algorithm	I	with the 1st (yellow)	helped simplify the meaning of
	and domatic number in @math time. Finally, we present a family of polynomial	1			evaluates the subset convolved		- Shows once again how the	the individual passages?
	space approximation algorithms that find a number between @math and	1			in O(		model fails to understand	
	@math in time @math .	1			Comment:		words used in unusual settings	
		1			- Copied the first tokens only		(e.g. mathematics)	
1		1	1	I	- No indication of MDS	I		
1		1	1	I	- Random word change noted	I		
1		1	1	I	(red)	I		
		1			1			

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
5068	(1):	Another bunch of related works	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
1	Convolutional neural networks have gained a remarkable success in computer	include hyper-parameter	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	vision. However, most usable network architectures are hand-crafted and	optimization @cite , meta-	0.0596 (prec)	0.0723 (prec)	0.0513 (prec)	0.0811 (prec)	0.0952 (prec)	0.0899 (prec)
	usually require expertise and elaborate design. In this paper, we provide a block-	learning @cite and learning to	0.1169 (recall)	0.0779 (recall)	0.1299 (recall)	0.039 (recall)	0.1299 (recall)	0.1039 (recall)
	wise network generation pipeline called BlockQNN which automatically builds	learn methods @cite @cite .	0.0789 (f-1)	0.075 (f-1)	0.0735 (f-1)	0.0526 (f-1)	0.1099 (f-1)	0.0964 (f-1)
	high-performance networks using the Q-Learning paradigm with epsilon-greedy	However, the goal of these works	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	exploration strategy. The optimal network block is constructed by the learning	is to use meta-data to improve	0.1053 (prec)	0.1548 (prec)	0.0969 (prec)	0.1842 (prec)	0.1698 (prec)	0.2111 (prec)
	agent which is trained sequentially to choose component layers. We stack the	the performance of the existing	0.2051 (recall)	0.1667 (recall)	0.2436 (recall)	0.0897 (recall)	0.2308 (recall)	0.2436 (recall)
	block to construct the whole auto-generated network. To accelerate the	algorithms, such as finding the	0.1391 (f-1)	0.1605 (f-1)	0.1387 (f-1)	0.1207 (f-1)	0.1957 (f-1)	0.2262 (f-1)
	generation process, we also propose a distributed asynchronous framework and	optimal learning rate of	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	an early stop strategy. The block-wise generation brings unique advantages: (1)	optimization methods or the	Convolutional neural networks	in this paper we show how the	Convolutional neural networks	In this paper, we propose a	Recently, meta-learning @cite	@cite @cite proposed a gradient
	it performs competitive results in comparison to the hand-crafted state-of-the-	optimal number of hidden layers	have gained a remarkable	design of an optimization	have gained a remarkable	block-wise network generation	@cite has gained a lot of	descent method for meta-
	art networks on image classification, additionally, the best network generated	to construct the network. In this	success in computer vision.	algorithm can be cast as a	success in computer vision.	pipeline using the Q-learning	attention in the machine learning	learning. The gradient descent
	by BlockQNN achieves 3.54 top-1 error rate on CIFAR-10 which beats all existing	paper, we focus on learning the	However, most usable network	learning problem, allowing the	However, most usable network	paradigm with epsilon-greedy	literature. In this paper, we	method is similar to our method,
	auto-generate networks. (2) in the meanwhile, it offers tremendous reduction	entire topological architecture of	architectures are hand-crafted	algorithm to learn to exploit	architectures are hand-crafted	exploration strategy. We also	propose a block-wise network	but differs from our method in
	of the search space in designing networks which only spends 3 days with 32	network blocks to improve the	and usually require expertise and	structure in the problems of	and usually require expertise and		generation pipeline called	two important aspects. First, the
	GPUs, and (3) moreover, it has strong generalizability that the network built on	performance.	elaborate design. In this paper,	interest in an automatic way.	elaborate design. In this paper,	asynchronous framework and an	BlockQNN which automatically	gradient descent method does
	CIFAR also performs well on a larger-scale ImageNet dataset.		we provide a block-wise network	we make meta-learning in large	we provide a block-wise network	early stop strategy to accelerate	builds high-performance	not rely on the gradient of the
1	(2):	1	generation pipeline called	systems feasible by using	generation pipeline called	the generation process.	networks using the Q-Learning	gradient. Second, it does not
	This paper introduces the application of gradient descent methods to meta-		BlockQNN which automatically	recurrent neural networks withth	BlockQNN which automatically	Comment:	paradigm with epsilon-greedy	require the gradient of a gradient
	learning. The concept of "meta-learning", i.e. of a system that improves or		builds high-performance	eir attendant learning routines as	builds high-performance	- Only extracted 2 sentences	exploration strategy. The optimal	to be learned. In this paper, we
	discovers a learning algorithm, has been of interest in machine learning for		networks using the Q-Learning	meta-learning systems.	networks using the Q-learning	from the 1st article (yellow)	network block is constructed	show that the gradient of our
	decades because of its appealing applications. Previous meta-learning		paradigm with epsilon-greedy	our learned algorithms	paradigm with epsilon-greedy	- No indication of MDS	sequentially by the learning	method can be used as a meta-
	approaches have been based on evolutionary methods and, therefore, have		exploration strategy. The optimal	outperform generic, hand-	exploration strategy. The optimal		agent which is trained	learning method. We also show
	been restricted to small models with few free parameters. We make meta-		network block is constructed by	designed competitors on the	network block is constructed by		sequentially to choose	that our approach can be applied
	learning in large systems feasible by using recurrent neural networks withth eir attendant learning routines as meta-learning systems. Our system derived		the learning agent which is	tasks for which they are trained, and also generalize well to new	the learning agent which is trained sequentially to choose		component layers. In addition, it offers tremendous reduction of	to large systems with few free parameters.
	complex well performing learning algorithms from scratch. In this paper we also		trained sequentially to choose component layers. We stack the	tasks with similar structure.	component lavers. We stack the		the search space in designing	Comment:
	show that our approachp erforms non-stationary time series prediction.		block to construct the whole	Comment:	block to construct the whole		networks which only spend 3	- Shows some sign of MDS, with
	(3):		auto-generated network. To	- Copied most of the 4th article	auto-generated network. To		days with 32 GPUs, and it has	the 2nd article (green) being
	Different researchers hold different views of what the term meta-learning		accelerate the generation	(purple) and just inserted a	accelerate the generation		strong generalizability that the	mentioned.
	exactly means. The first part of this paper provides our own perspective view in		process, we also propose a	sentence from the 2nd (green)	process, we also propose a		network built on CIFAR-10 also	- However, the other parts are
	which the goal is to build self-adaptive learners (i.e. learning algorithms that		distributed asynchronous	- No indication of MDS	distributed asynchronous		performs well on a larger-scale	mostly hallucinations
	improve their bias dynamically through experience by accumulating meta-		framework and an early stop		framework and an early stop		ImageNet dataset.	
	knowledge). The second part provides a survey of meta-learning as reported by		strategy. The block-wise		strategy. The block-wise		Comment:	
	the machine-learning literature. We find that, despite different views and		generation brings unique		generation brings unique		- No indication of MDS, except	
	research lines, a question remains constant: how can we exploit knowledge		advantages: (1) it performs		advantages: (1) it performs		for the red part where the	
	about learning (i.e. meta-knowledge) to improve the performance of learning		competitive results in		competitive results in		model noted "meta-learning"	
	algorithms? Clearly the answer to this question is key to the advancement of		comparison to the hand-crafted		comparison to the hand-crafted		is a recurring theme	
	the field and continues being the subject of intensive research.		state-of-the-art networks on		state-of-the-art networks on			
	(4):		image classification, additionally,		image classification, additionally,			
	The move from hand-designed features to learned features in machine learning		the best network generated by		the best network generated by			
	has been wildly successful. In spite of this, optimization algorithms are still		BlockQNN achieves 3.54 top-1		BlockQNN achieves 3.54 top-1			
	designed by hand. In this paper we show how the design of an optimization		error rate on CIFAR-10 which		error rate on CIFAR-10 which			
	algorithm can be cast as a learning problem, allowing the algorithm to learn to		beats all existing auto-generate		beats all existing auto-generate			
	exploit structure in the problems of interest in an automatic way. Our learned		networks. (2)		networks. (2) in the meanwhile,			
	algorithms, implemented by LSTMs, outperform generic, hand-designed		Comment:		it offers tremendous reduction of			
	competitors on the tasks for which they are trained, and also generalize well to		- Copied the first tokens only		the search space in designing			
	new tasks with similar structure. We demonstrate this on a number of tasks,		- No indication of MDS		networks which only spends 3			
	including simple convex problems, training neural networks, and styling images				days with 32 GPUs, and (3)			
1	with neural art.	1			moreover, it has strong			
1		1			generalizability that the network			
1		1			built on CIFar also performs well			
1		1			on a larger-scale ImageNet dataset.			
1		1			Comment:			
1		1			- Copied the first tokens only			
		1			- No indication of MDS			
1		1						
1		1						
		1			l .	l .		

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
76	(1):	Representing text Early works in	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	We address the problem of cross-modal fine-grained action retrieval between	image-to-text cross-modal	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	text and video. Cross-modal retrieval is commonly achieved through learning a	retrieval @cite @cite used	0.0194 (prec)	0.0327 (prec)	0.0279 (prec)	0.0458 (prec)	0.0272 (prec)	0.0455 (prec)
	shared embedding space, that can indifferently embed modalities. In this	TF-IDF as a weighted bag-of-	0.0319 (recall)	0.0532 (recall)	0.0532 (recall)	0.0638 (recall)	0.0426 (recall)	0.0745 (recall)
	paper, we propose to enrich the embedding by disentangling parts-of-speech	words model for text	0.0241 (f-1)	0.0405 (f-1)	0.0366 (f-1)	0.0533 (f-1)	0.0332 (f-1)	0.0565 (f-1)
	(PoS) in the accompanying captions. We build a separate multi-modal	representations (either from a	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	embedding space for each PoS tag. The outputs of multiple PoS embeddings	word embedding model or one-	0.0769 (prec)	0.1169 (prec)	0.1056 (prec)	0.1439 (prec)	0.1554 (prec)	0.1097 (prec)
	are then used as input to an integrated multi-modal space, where we perform	hot vectors) in order to aggregate	0.1263 (recall)	0.1895 (recall)	0.2 (recall) 0.1382 (f-1)	0.2 (recall)	0.2421 (recall)	0.1789 (recall)
	action retrieval. All embeddings are trained jointly through a combination of PoS-aware and PoS-agnostic losses. Our proposal enables learning specialised	variable length text captions into a single fixed sized	0.0956 (f-1) Summary:	0.1446 (f-1) Summary:	0.1382 (f-1) Summarv:	0.1674 (f-1) Summary:	0.1893 (f-1) Summary:	0.136 (f-1) Summary:
	embedding spaces that offer multiple views of the same embedded entities.	representation. With the advent	-modal features (different visual	we address the problem of	"We address the problem of	In the context of video-to-text	Cross-modal retrieval has been a	In @cite @cite, the authors
	We report the first retrieval results on fine-grained actions for the large-scale	of neural networks, works shifted	characteristics, audio inputs, and	cross-modal fine-grained action	cross-modal fine-grained action	retrieval, the task of image-to-	long-standing research problem	propose to use a PoS-agnostic
	EPIC dataset, in a generalised zero-shot setting. Results show the advantage of	to use RNNs, Gated Recurrent	text) by a fusion strategy for	retrieval between text and	retrieval between text and	vided retrieval has been studied	in computer vision and natural	loss to enrich the embedding
	our approach for both video-to-text and text-to-video action retrieval. We also	Units (GRU) or Long Short-Term	efficient retrieval. Furthermore,	video.	video. Cross-modal retrieval is	for decades @cite @cite, @cite	language processing (NLP) @cite	space by disentangling the parts-
	demonstrate the benefit of disentangling the PoS for the generic task of cross-	Memory (LSTM) units to extract	we explore several loss functions	we propose a novel framework	commonly achieved through	and @cite. In recent years, a	@cite. In recent years, several	of-speech (PoS) in the
	modal video retrieval on the MSR-VTT dataset.	textual features @cite or to use	in training the embedding and	for learning joint embeddings of	learning a shared embedding	'number of approaches have	approaches have been proposed	accompanying captions.
	(2):	these models within the	propose a modified pairwise	images and text using a two-	space, that can indifferently	been proposed to learn joint	to learn joint embeddings for	However, the authors do not
	We describe a novel cross-modal embedding space for actions, named	embedding network @cite @cite	ranking loss for the task.	branch neural network with	embed modalities. In this paper,	embeddings of images and text	image-text and text-to-video	consider the use of the PoS loss
	Action2Vec, which combines linguistic cues from class labels with spatio-	@cite @cite @cite for both	Experiments on MSVD and MSR-	multiple layers of linear	we propose to enrich the	using a two-branch neural	tasks. @cite proposed to learn a	in the context of video-to-video
	temporal features derived from video clips. Our approach uses a hierarchical	modalities.	VTT datasets demonstrate that	projections followed by	embedding by disentangling	network with multiple layers of	joint embedding of visual and	action retrieval. In @cite, the
	recurrent network to capture the temporal structure of video features. We		our method achieves significant	nonlinearities.	parts-of-speech (PoS) in the	linear projections followed by	textual cues that allows one to	authors propose a joint
1	train our embedding using a joint loss that combines classification accuracy		performance gain compared to	the network is trained using a	accompanying captions. We	nonlinearities. @cite proposed a	query the database using a text	embedding space for noun and
	with similarity to Word2Vec semantics. We evaluate Action2Vec by performing		the state-of-the-art	largemargin objective that	build a separate multi-modal	multi-modal embedding for	modifier in addition to the query	action videos by augmenting the
	zero shot action recognition and obtain state of the art results on three		approaches.    Learning a	combines cross-view ranking	embedding space for each PoS	image and text embeddings,	image, adapting the results to	training set with millions of
1	standard datasets. In addition, we present two novel analogy tests which		joint language-visual embedding	constraints with within-view	tag. The outputs of multiple PoS	which can be trained jointly	the modifier. In @cite, a multi-	weak annotated images with
	quantify the extent to which our joint embedding captures distributional semantics. This is the first joint embedding space to combine verbs and action		has a number of very appealing properties and can result in	neighborhood structure preservation constraints inspired	embeddings are then used as input to an integrated multi-	through a combination of neural language models and	modal joint embedding space is proposed to capture the	strong and noisy annotations @cite. @cite proposes to use a
1	videos, and the first to be thoroughly evaluated with respect to its		variety of practical application,	by metric learning literature.	Modal space, where we perform	multimodal neural language	temporal structure of video	two-branch neural network with
	distributional semantics.		including natural language	extensive experiments show	action retrieval. All	models. In @cite a joint	features. The joint embedding is	multiple layers of linear
	(3):		image video annotation and	that our approach gains	embeddeddings are trained	embedding for images and text	then used to disambiguate fine-	projections followed by
	Inspired by recent advances in multimodal learning and machine translation,		search. In this work, we study	significant improvements in	jointly through a combination of	embedding was proposed for	grained visual concepts in the	nonlinearities. The network is
	we introduce an encoder-decoder pipeline that learns (a): a multimodal joint		three different joint language-	accuracy for image-to-text and	PoS-aware and PoS-agnostic	image-sentence retrieval, where	sentence embedding process. In	trained using a largemargin
	embedding space with images and text and (b): a novel language model for		Visual neural network model	textto-image retrieval.	losses. Our proposal enables	the goal is to retrieve images	addition, @cite and @cite	objective that combines cross-
	decoding distributed representations from our space. Our pipeline effectively		architectures. We evaluate our	in addition, we show that a	learning specialised embedding	that contain the same object	propose to learn an embedding	view ranking constraints with
	unifies joint image-text embedding models with multimodal neural language		models on large scale LSMDC16	similarity based on human-	spaces that offer multiple views	instance as the query image. In	of images where the similarity in	within-view neighborhood
	models. We introduce the structure-content neural language model that		movie dataset for two tasks: 1)	annotated region-level captions	of the same embedded entities.	this work, we focus on the task	the visual space is correlated	structure preservation
	disentangles the structure of a sentence to its content, conditioned on		Standard Ranking for video	is highly correlated with the	We report the first retrieval	of video and text retrieval.	with the semantic similarity	constraints inspired by metric
	representations produced by the encoder. The encoder allows one to rank		annotation and retrieval 2) Our proposed movie multiple-choice	human ranking and constitutes a good computable surrogate.	results on fine-grains actions for the large-scale EPIC dataset, in a	Comment: - Some signs of MDS, with the	surrogate. In @math, @cite learn a joint representation	learning literature. However,
	scratch. Using LSTM to encode sentences, we match the state-of-the-art		proposed movie multiple-choice test. This test facilitate	we further extend our model to	generalised zero-shot setting.	- Some signs of MDS, with the 1st (yellow) article used as	invariant across different video	their method is based on a multi-layer neural network with
	performance on Flickr8K and Flickr30K without using object detections. We		automatic evaluation of visual-	learn a joint embedding of	Results show the advantage of	context, and the 4th (purple)	modalities	a single layer of linear
	also set new best results when using the 19-layer Oxford convolutional		language models for natural	images and textual cues that	our approach for both video-to-	and 8th (deep blue) articles	Comment:	projections and nonlinearities
	network. Furthermore we show that with linear encoders, the learned		language video annotation	allows one to query the	text and text-to-video action	covered.	- Much stronger signs of MDS,	and is trained on the Flickr30K
	embedding space captures multimodal regularities in terms of vector space		based on human activities. In	database using a text modifier in	retrieval. We also demonstrate	- However, the 2 <sup>nd</sup> yellow part	with the information from the	Entities image-sentence dataset
	arithmetic e.g. *image of a blue car* - "blue" + "red" is near images of red		addition to original Audio	addition to the query image	the benefit of disentangling the	also shows that the work being	1st article (yellow) used as	@cite.
	cars. Sample captions generated for 800 images are made available for		Description (AD) captions	modifier.	PoS for the generic task of cross-	done in the main article is	context, and then the 2nd	Comment:
	comparison.		Comment:	finally, we show how the	Modal video retrieval on the	mistaken as one of the	(green), 4th (purple), 5th	- Shows some sign of MDS, with
	(4):		- Copied from the middle of the	proposed	MSR-VTT dataset." "We	references.	(orange), 7th (deep yellow) and	the multiple articles being
	Querying with an example image is a simple and intuitive interface to retrieve		5th article (orange)	Comment:	Comment:	- The last part on the task of the	8th (deep blue) all referenced.	mentioned.
1	information from a visual database. Most of the research in image retrieval has		- No sign of MDS	- Very weak signs of MDS, with	- Copied the first tokens only	main paper is possibly a		- However, the red parts are
	focused on the task of instance-level image retrieval, where the goal is to			the 1st (yellow), 4th (purple)	- No indication of MDS	hallucination.		mostly hallucinations
1	retrieve images that contain the same object instance as the query image. In this work we move beyond instance-level retrieval and consider the task of			and 8th (deep blue) articles covered.				
	semantic image retrieval in complex scenes, where the goal is to retrieve			- However, they are all				
	images that share the same semantics as the query image. We show that,			presented as a single				
	despite its subjective nature, the task of semantically ranking visual scenes is			document.				
	consistently implemented across a pool of human annotators. We also show			- The model also performs like				
	that a similarity based on human-annotated region-level captions is highly			an extractive model				
1	correlated with the human ranking and constitutes a good computable							
	surrogate. Following this observation, we learn a visual embedding of the							
	images where the similarity in the visual space is correlated with their							
	semantic similarity surrogate. We further extend our model to learn a joint							
1	embedding of visual and textual cues that allows one to query the database				I	l		
	using a text modifier in addition to the query image, adapting the results to							
	the modifier. Finally, our model can ground the ranking decisions by showing regions that contributed the most to the similarity between pairs of images,							
	providing a visual explanation of the similarity.							
	(5):							
1	Constructing a joint representation invariant across different modalities (e.g.,				I	l		
	video, language) is of significant importance in many multimedia applications.							
1	While there are a number of recent successes in developing effective image-				I	l		
	text retrieval methods by learning joint representations, the video-text							
1	retrieval task, however, has not been explored to its fullest extent. In this							
1	paper, we study how to effectively utilize available multimodal cues from							
	videos for the cross-modal video-text retrieval task. Based on our analysis, we							

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
485	(1):	In learning classifier with web	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Learning from web data has attracted lots of research interest in recent years.	data, previous works focus on	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	However, crawled web images usually have two types of noises, label noise	handling the label noise in three	0.0175 (prec)	0.0258 (prec)	0.0048 (prec)	0.0485 (prec)	0.0266 (prec)	0.0287 (prec)
	and background noise, which induce extra difficulties in utilizing them	directions, removing label noise	0.0732 (recall)	0.0976 (recall)	0.0244 (recall)	0.122 (recall)	0.122 (recall)	0.1463 (recall)
	effectively. Most existing methods either rely on human supervision or ignore	@cite @cite @cite @cite	0.0283 (f-1)	0.0408 (f-1)	0.0081 (f-1)	0.0694 (f-1)	0.0437 (f-1)	0.048 (f-1)
	the background noise. In this paper, we propose a novel method, which is	@cite @cite , building noise-	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	capable of handling these two types of noises together, without the	robust model @cite @cite @cite	0.064 (prec)	0.0705 (prec)	0.0433 (prec)	0.1058 (prec)	0.0688 (prec)	0.0667 (prec)
	supervision of clean images in the training stage. Particularly, we formulate our	@cite @cite @cite , and	0.2619 (recall)	0.2619 (recall)	0.2143 (recall)	0.2619 (recall)	0.3095 (recall)	0.3333 (recall)
	method under the framework of multi-instance learning by grouping ROIs (i.e.,	curriculum learning @cite @cite .	0.1028 (f-1)	0.1111 (f-1)	0.072 (f-1)	0.1507 (f-1)	0.1126 (f-1)	0.1111 (f-1)
	images and their region proposals) from the same category into bags. ROIs in		Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	each bag are assigned with different weights based on the representative		Learning from web data has	learning from web data is	We present a simple yet efficient	Label noise is an important issue	The label noise problem has	Recently, there has been a lot of
	discriminative scores of their nearest clusters, in which the clusters and their		attracted lots of research	increasingly popular due to	approach capable of training	in image classification, with	been widely studied in the	research on label noise cleaning.
	scores are obtained via our designed memory module. Our memory module		interest in recent years.	abundant free web resources.	deep neural networks on large-	many potential negative	literature @cite @cite. In @cite,	In @cite @cite, the label noise is
	could be naturally integrated with the classification module, leading to an end-		However, it is still very difficult	however, the performance gap	scale weakly-supervised web	consequences @cite @cite. For	the label noise is treated as a	treated as a set of instances that
	to-end trainable system. Extensive experiments on four benchmark datasets		to learn from web data without	between webly supervised	images, which are crawled raw	example, the label noise can be	label noise, and the label noise	are mislabeled, and the label
	demonstrate the effectiveness of our method.		the supervision of a human. In	learning and traditional	from the Internet by using text	class-conditional @cite, which is	can be class-conditional. @cite	cleansing algorithm is proposed
	(2):		this paper, we study the problem	supervised learning is still very	queries, without any human	not scalable for large-scale	and @cite address the label	to deal with the label noise. In
	We study the problem of automatically removing outliers from noisy data,		of learning from web data. The	large, due to the label noise of	supervision. We develop a	weakly-supervised learning. To	noise by using the abundant	addition, @cite proposed an
	with application for removing outlier images from an image collection. We		problem is that web images	web data as well as the domain	principled learning strategy by	address this problem, several	surrogate loss functions	active learning approach for
	address this problem by utilizing the reconstruction errors of an autoencoder.		usually have two types of noises,	shift between web data and test	leveraging curriculum learning,	methods have been proposed.	designed for the traditional	removing outliers from an image
	We observe that when data are reconstructed from low-dimensional		label noise and background	data. to fill this gap, most	with the goal of handling a	@cite proposed a method for	classification problem when	collection @cite, which does not
	representations, the inliers and the outliers can be well separated according to		noise, which induce extra	existing methods propose to	massive amount of noisy labels	earning from unlabeled web	there is label noise. Recently,	require a large amount of
	their reconstruction errors. Based on this basic observation, we gradually		difficulties in utilizing them	purify or augment web data	and data imbalance effectively.	data, which is based on the	@cite proposed a method to	training data to train the model,
	inject discriminative information in the learning process of an autoencoder to		effectively. Most existing	using instance-level supervision,	We design a new learning	assumption that the unlabeled	remove outlier images from an	and they do not need any
	make the inliers and the outliers more separable. Experiments on a variety of		methods either rely on human	which generally requires heavy	curriculum by measuring the	data can be used to improve the	image collection by utilizing the	annotated labels for training
	image datasets validate our approach. (3):		supervision or ignore the	annotation. instead, we propose to address	complexity of data using its	classification performance.  However, these methods do not	reconstruction errors of an autoencoder. In this paper, we	Moreover, the label noise cleansing algorithms are defined
			background noise. In this article,		distribution density in a feature			
	We present a theoretically grounded approach to train deep neural networks, including recurrent networks, subject to class-dependent label noise. We		we propose a novel method, which is capable of handling	the label noise and domain shift by using more accessible	space, and rank the complexity in an unsupervised manner. This	consider the label noise of web data. In contrast, our method is	propose a novel method, which is capable of handling both label	as follows: first, the labels of the instances are labeled, and
	propose two procedures for loss correction that are agnostic to both		these two types of noise	category-level supervision.	allows for an efficient	able to deal with the label noise	noise and background noise	second, the labels on the labels
	application domain and network architecture. They simply amount to at most		together, without the	in particular, we build our deep	implementation of curriculum	without the supervision of the	together, without the	are labeled, respectively. The
	a matrix inversion and multiplication, provided that we know the probability of		supervision or supervision of	probabilistic framework upon	learning on large- scale web	classification module	supervision of clean images in	labels are labeled by the label
	each class being corrupted into another. We further show how one can		clean images in the training	variational autoencoder (VAE ),	images, resulting in a high-	Comment:	the training stage. In addition,	noise correction algorithm and
	estimate these probabilities, adapting a recent technique for noise estimation		stage. Particularly, we formulate	in which classification network	performance CNN the model,	- Some sign of MDS, using the	we formulate our method under	the labels of mislabeled
	to the multi-class setting, and thus providing an end-to-end framework.		our method under the	and VAE can jointly leverage	where the negative impact of	context information from the	the framework of multi-instance	instances are labeled by label
	Extensive experiments on MNIST, IMDB, CIFAR-10, CIFAR-100 and a large scale		framework of multi-instance	category-level hybrid	noisy labels is reduced	1st article (yellow) as	learning by grouping ROIs (i.e.,	noise cleansing. @cite used
	dataset of clothing images employing a diversity of architectures —		learning by grouping ROIs (i.e.,	information, our memory	substantially. Importantly, we	background, and also	images and their region	curriculum learning to train a
	stacking dense, convolutional, pooling, dropout, batch normalization, word		images and their region	module could be naturally	show by experiments that those	mentioned correctly some	proposals) from the same	deep neural network on large-
	embedding, LSTM and residual layers — demonstrate the noise		proposals) from the same	integrated with the classification	images with highly noisy labels	elements of the current study	category into bags. ROIs in each	scale web images, which are
	robustness of our proposals. Incidentally, we also prove that, when ReLU is the		category into bags. ROIs in each	module, leading to an end-to-	can surprisingly improve the	- Elements of 4th article (purple)	bag are assigned with different	crawled raw from the Internet
	only non-linearity, the loss curvature is immune to class-dependent label		bag are assigned with different	end trainable system.	generalization capability of	is clear while the 1st red part is	weights based on the	by using text queries, without
	noise.		weights based on the	we design a new learning	model, by serving as a manner	true for many samples.	representative discriminative	any human annotation.
	(4):		representative discriminative	curriculum by measuring the	of regularization. Our		scores of their nearest clusters,	However, curriculum learning
	In this paper, we study the problem of learning image classification models		scores of their nearest clusters,	complexity of data using its	approaches obtain state-of-the-		in which the clusters and their	can significantly improve the
	with label noise. Existing approaches depending on human supervision are		in which the clusters and their	distribution density, in an	art performance on four		scores are obtained via our	generalization capability of the
	generally not scalable as manually identifying correct or incorrect labels is		scores are obtained via our	unsupervised manner in an	benchmarks: WebVision,		designed memory module. Our	model, by reducing the number
	time-consuming, whereas approaches not relying on human supervision are		designed memory module.	Comment:	ImageNet, Clothing-1M and		memory module could be	of noisy labels, and by using
	scalable but less effective. To reduce the amount of human supervision for		Comment:	- Copied the 6th article, except	Food-101. With an ensemble of		naturally integrated with the	curriculum learning to reduce
	label noise cleaning, we introduce CleanNet, a joint neural embedding		- Copied the first tokens only	replacing the last 2 sentences			classification module, leading to	the amount of data imbalance in
	network, which only requires a fraction of the classes being manually verified		- No indication of MDS	with another sentence from	top-5 error rate of 5.2 on the		an end-to-end trainable system.	the training process. Moreover,
	to provide the knowledge of label noise that can be transferred to other			the 5th article (orange)	WebVision challenge [18] for		Comment:	curriculum learning is effective
	classes. We further integrate CleanNet and conventional convolutional neural			- No indication of MDS	1000-category classification. This		<ul> <li>Clear signs of MDS, with the 1st</li> </ul>	in handling noisy labels and the
	network classifier into one framework for image classification learning. We				result was the top performance		article (yellow) used for both	data imbalance effectively.
	demonstrate the effectiveness of the proposed algorithm on both of the label				by a wide margin,		context (at the beginning) and	comment.
	noise detection task and the image classification on noisy data task on several				outperforming second place by a		contrast with other studies in the latter half	- Some sign of MDS, with the 1st
1	large-scale datasets. Experimental results show that CleanNet can reduce label noise detection error rate on held-out classes where no human supervision				nearly 50 relative	l	the latter half  - 3 other articles are cited,	article (yellow) used for context, and the 2 <sup>nd</sup> (green),
	noise detection error rate on held-out classes where no human supervision available by 41.5 compared to current weakly supervised methods. It also				- Copied the 5th article only		<ul> <li>3 other articles are cited, though the red parts shows a</li> </ul>	7th (pink) and 8th (deep blue)
	achieves 47 of the performance gain of verifying all images with only 3.2				- No indication of MDS		poorly written sentence.	articles referenced.
	images verified on an image classification task. Source code and dataset will be				- NO IIIUICALIOII OI WIDS		poorly written sentence.	- Contains much hallucinations
	available at kuanghuei.github.io CleanNetProject.							though as noted in the red
	(5):							parts.
	We present a simple yet efficient approach capable of training deep neural							pur co.
	networks on large-scale weakly-supervised web images, which are crawled raw							
	from the Internet by using text queries, without any human annotation. We							
	develop a principled learning strategy by leveraging curriculum learning, with							
	the goal of handling a massive amount of noisy labels and data imbalance							
	effectively. We design a new learning curriculum by measuring the complexity							
	of data using its distribution density in a feature space, and rank the							
	complexity in an unsupervised manner. This allows for an efficient							
	implementation of curriculum learning on large-scale web images, resulting in							
	a high-performance CNN the model, where the negative impact of noisy labels							
	is reduced substantially. Importantly, we show by experiments that those							
	images with highly noisy labels can surprisingly improve the generalization							
	capability of model, by serving as a manner of regularization. Our approaches							

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
1368	(1):	Profiling plays a key role in	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	Low latency services such as credit-card fraud detection and website targeted	managed runtimes, either for	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	advertisement rely on Big Data platforms (e.g., Lucene, Graphchi, Cassandra)	code optimization or memory	0.0323 (prec)	0.0 (prec)	0.0142 (prec)	0.0482 (prec)	0.0169 (prec)	0.0164 (prec)
	which run on top of memory managed runtimes, such as the JVM. These	management decisions @cite	0.037 (recall)	0.0 (recall)	0.0222 (recall)	0.0296 (recall)	0.0074 (recall)	0.0148 (recall)
	platforms, however, suffer from unpredictable and unacceptably high pause	@cite @cite @cite @cite	0.0345 (f-1)	0.0 (f-1)	0.0173 (f-1)	0.0367 (f-1)	0.0103 (f-1)	0.0156 (f-1)
	times due to inadequate memory management decisions (e.g., allocating	@cite . We focus on getting	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	objects with very different lifetimes next to each other, resulting in memory	quality profiling information to	0.1026 (prec)	0.12 (prec)	0.1038 (prec)	0.1786 (prec)	0.2 (prec)	0.1301 (prec)
	fragmentation). This leads to long and frequent application pause times,	drive object pretenuring. is, to	0.1176 (recall)	0.1324 (recall)	0.1618 (recall)	0.1103 (recall)	0.0882 (recall)	0.1176 (recall)
	breaking Service Level Agreements (SLAs). This problem has been previously identified and results show that current memory management techniques are	the best of our knowledge, the first online profiler targeting the	0.1096 (f-1) Summary:	0.1259 (f-1) Summary:	0.1264 (f-1) Summary:	0.1364 (f-1) Summary:	0.1224 (f-1) Summary:	0.1236 (f-1) Summary:
	ill-suited for applications that hold in memory massive amounts of middle to	dynamic pretenuring of objects in	Low latency services such as	this paper presents the	Programmers are writing a	Object lifetime profiling @cite	@cite @cite propose a tunable	@cite @cite is a dynamic
	long-lived objects (which is the case for a wide spectrum of Big Data	Big Data applications running on	credit-card fraud detection and	architecture of the Jalapeno	rapidly growing number of	@cite has been widely studied in	profiling technique, called	program analysis tool for
	applications). Previous works try to reduce such application pauses by	HotSpot. This section compares	website targeted advertisement	Adaptive Optimization System, a	programs in object-oriented	the literature. However, most of	Resurrector, that explores the	producing traces of garbage
	allocating objects off-heap or in special allocation regions generations, thus	our work with state-of-art	rely on Big Data platforms (e.g.,	system to support leading-edge	languages, such as Java and C#,	the existing works focus on the	middle ground between high	collection events, including
	alleviating the pressure on memory management. However, all these solutions	systems, namely, off-line and	Lucene, Graphchi, Cassandra)	virtual machine technology and	that require garbage collection.	analysis of allocation contexts,	precision and high efficiency to	object allocations, object deaths,
	require a combination of programmer effort and knowledge, source code	online profilers that guide	which run on top of memory	enable ongoing research on	Garbage collection traces and	such as stack allocation @cite,	find the precision-efficiency	and pointer updates. It is based
	access, or off-line profiling, with clear negative impact on programmer	systems where small changes are	managed runtimes, such as the	online feedback-directed	simulation speed up research by	object allocations @cite and	sweetspot for various liveness-	on the Merlin algorithm @cite,
	productivity and or application performance. This paper presents ROLP, a	needed in the heap organization	JVM. These platforms, however,	optimizations.	enabling deeper understandings	pointer updates @cite. These	based optimization techniques.	which is based on Java
	runtime object lifetime profiling system. ROLP profiles application code at	and collection. It ends with a	suffer from unpredictable and	we describe the extensible	of object lifetime behavior and	approaches are based on the	@cite present an	Reference Objects (JCU) and is
	runtime in order to identify which allocation contexts create objects with	comparative analysis of systems	unacceptably high pause times.	system architecture, based on a	quick exploration and design of	Merlin algorithm @cite which is	implementation of the Jalapeno	designed to be used with any
	middle to long lifetimes, given that such objects need to be handled differently	that demand a more profound	Merlin timestamps objects and	federation of threads with	new garbage collection	based on the JVM Tool Interface	Adaptive Optimization System, a	standard JVM. It is not designed
	(regarding short-lived ones). This profiling information greatly improves	change, either to the application	later uses the timestamps of	asynchronous communication, and we present an	algorithms. When generating	(JVMTI) Microsystems @cite that	system to support leading-edge	for dynamic program analysis,
	memory management decisions, leading to long tail latencies reduction of up to 51 for Lucene, 85 for GraphChi, and 60 for Cassandra, with negligible	framework or the runtime itself,	dead objects to reconstruct when they died. The Merlin	and we present an implementation of the general	perfect traces, the brute-force method of computing object	supports adaptive multi-level optimization based on statistical	virtual machine technology and	since it does not support dynamic program analysis and
	throughput and memory overhead. ROLP is implemented for the OpenJDK 8	in some cases manipulating application-defined types and or	algorithm piggybacks on garbage	architecture that supports	lifetimes requires a whole-heap	sampling. However, these	enable ongoing research on online feedback-directed	does not consider memory
	HotSpot JVM and it does not require any programmer effort or source code	organizing the heap in special	collections performed by the	adaptive multi-level optimization	garbage collection at every	approaches are not suitable for	optimizations.	usage, which makes it not
	access.	purpose regions, and placing data	base system. Experimental	based purely on statistical an	potential collection point in the	large-scale applications, such as	Comment:	suitable for dynamic analysis,
	(2):	directly in an off-heap space.	results show that Merlin can	online feedback - directed	program. Because this process is	data structures with disjoint	- Slightly weaker MDS	
	Future high-performance virtual machines will improve performance through		generate traces over two orders	inlining optimization based on	prohibitively expensive.	lifetimes.	performance compared to	implementations. @cite present
	sophisticated online feedback-directed optimizations. this paper presents the		of magnitude faster than the	statistical edge sampling	researchers often use granulated	Comment:	tuned LED, with the 2nd (green)	a technique to make any profiler
	architecture of the Jalapeno Adaptive Optimization System, a system to		brute-force method which	sampling.	traces by collecting only	- Clear signs of MDS, with the 1st	and 6th (orange) articles	aware of optimizations
	support leading-edge virtual machine technology and enable ongoing research		collects after every object	the paper also describes and	periodically, for example, every	article (yellow) used for	referenced	performed by the dynamic
	on online feedback-directed optimizations. We describe the extensible system		allocation. We also use Merlin to	evaluates an online feedback -	32 KB of allocation. We extend	context (at the beginning),	<ul> <li>Compared to the LED summary,</li> </ul>	compiler. They demonstrate that
	architecture, based on a federation of threads with asynchronous		produce visualizations of heap	directed inlining optimization	the state of the art for	with the 3rd (blue), 4th (purple)	no reference is made as to	their approach can be used to
	communication. We present an implementation of the general architecture		behavior that expose new object	based on statistical edge	simulating garbage collection	and 5th (deep blue) articles are	what the main (i.e. 1st articles)	detect errors in dynamic
	that supports adaptive multi-level optimization based purely on statistical		lifetime behaviors.    Modern object-oriented applications	sampling. this paper presents a tunable	algorithms in two ways. First, we develop a systematic	reference The last yellow sentence is in	does	compiler implementations, and
	sampling. We empirically demonstrate that this profiling technique has low overhead and can improve startup and steady-state performance, even		commonly suffer from severe	profiling technique, called	methodology for simulation	line with the 1st article, though		it does not rely on escape analysis.
	without the presence of online feedback-directed optimizations. The paper		performance problems that	Resurrector, that explores the	studies of garbage collection and	the words in red are adopted		Comment:
	also describes and evaluates an online feedback-directed inlining optimization		need to be optimized away for	middle ground between high	present results showing the	from the 6th article which the		- Weakest of the three tuned
	based on statistical edge sampling. The system is written completely in Java,		increased efficiency and user	precision and high efficiency to	effects of trace granularity on	commenter does not have		models but still shows MDS
	applying the described techniques not only to application code and standard		satisfaction. Many existing	find the precision-efficiency	these simulations. We show that	knowledge to judge.		signs, with the 3rd (blue) and
	libraries, but also to the virtual machine itself.		optimization techniques (such as	sweetspot for various	trace granularity often distorts			4th (purple) articles
	(3):		object pooling and pretenuring)	livenessbased optimization	simulated garbage collection			summarized
	We introduce Elephant Tracks (ET), a dynamic program analysis tool for		require precise identification of	techniques.	results compared with perfect			- Red part is basically
	producing traces of garbage collection events, including object allocations,		object lifetimes. However, all	we empirically demonstrate	traces. Second, we present and			hallucination though
	object deaths, and pointer updates. Like prior work, our tracing tool is based		these	that this profiling technique has	measure the performance of a			
	on the Merlin algorithm [2002], but offers several substantial new capabilities.		Comment:	low overhead and can improve	new algorithm called Merlin for			
	First, it is much more precise than previous tools: it measures time in terms of method entries and exits, allowing it to place events precisely in the context of		Copied the first tokens only     No indication of MDS	startup and steady-state performance, even without the	computing objects lifetimes.  Merlin timestamps objects and			
	the program structure. Furthermore, we exploit this notion of time to actually		- NO Indication of MD3	presence of online feedback-	later uses the timestamps of			
	reduce the amount of work required to compute precise object death times.			directions.	dead objects to reconstruct			
	Second, it is implemented using a combination of JVM Tool Interface (JVMTI)			our approach enables new kinds	when they died. The Merlin			
	Microsystems [2004] callbacks and bytecode rewriting, and works with any			of profilers, such as a profiler for	algorithm piggybacks on garbage			
	standard JVM. Finally, it properly handles Java Reference objects, including			Comment:	collections performed by the			
	weak references, finalization, and multithreaded programs.			- Copied the 2nd article only with	base system. Experimental			
	(4):			a reshuffling of sentence order	results show that Merlin can			
	Many profilers based on bytecode instrumentation yield wrong results in the			- No indication of MDS	generate traces over two orders			
	presence of an optimizing dynamic compiler, either due to not being aware of				of magnitude faster than the			
	optimizations such as stack allocation and method inlining, or due to the				brute-			
	inserted code disrupting such optimizations. To avoid such perturbations, we				Comment:			
	present a novel technique to make any profiler implemented at the bytecode level aware of optimizations performed by the dynamic compiler. We				Copied the 5th article only     No indication of MDS			
	implement our approach in a state-of-the-art Java virtual machine and		1		- NO INDICACION OF WIDS			
	demonstrate its significance with concrete profilers. We quantify the impact of							
	escape analysis on allocation profiling, object life-time analysis, and the impact							
	of method inlining on callsite profiling. We illustrate how our approach							
	enables new kinds of profilers, such as a profiler for non-inlined callsites, and a							
	testing framework for locating performance bugs in dynamic compiler							
	implementations.							
	(5):							
	Programmers are writing a rapidly growing number of programs in object-							
	oriented languages, such as Java and C#, that require garbage collection.							
	Garbage collection traces and simulation speed up research by enabling							
	deeper understandings of object lifetime behavior and quick exploration and							

W of in	t): While word embeddings have been shown to implicitly encode various forms	Several authors have tried to						
W. of in	(hiloord omboddings have been shown to implicitly enough years of forms		Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
of in ac		improve word embeddings by	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
ac	f attributional knowledge, the extent to which they capture relational	incorporating external knowledge	0.0533 (prec)	0.0284 (prec)	0.0333 (prec)	0.0633 (prec)	0.0795 (prec)	0.033 (prec)
	formation is far more limited. In previous work, this limitation has been	bases. For example, some	0.0616 (recall)	0.0274 (recall)	0.0479 (recall)	0.0342 (recall)	0.0479 (recall)	0.0205 (recall)
	ddressed by incorporating relational knowledge from external knowledge	authors have proposed models	0.0571 (f-1)	0.0279 (f-1)	0.0393 (f-1)	0.0444 (f-1)	0.0598 (f-1)	0.0253 (f-1)
	ases when learning the word embedding. Such strategies may not be	which combine the loss function	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	ptimal, however, as they are limited by the coverage of available resources	of a word embedding model, to	0.1588 (prec)	0.1549 (prec)	0.1327 (prec)	0.2375 (prec)	0.2472 (prec)	0.1848 (prec)
	nd conflate similarity with other forms of relatedness. As an alternative, in	ensure that word vectors are predictive of their context words,	0.1837 (recall) 0.1703 (f-1)	0.1497 (recall) 0.1522 (f-1)	0.1905 (recall) 0.1564 (f-1)	0.1293 (recall) 0.1674 (f-1)	0.1497 (recall)	0.1156 (recall) 0.1423 (f-1)
	mbedding, which is aimed to be complementary to a given standard word	with the loss function of a	0.1703 (T-1) Summary:	0.1522 (T-1) Summary:	0.1564 (T-1) Summary:	0.16/4 (f-1) Summary:	0.1864 (f-1) Summary:	0.1423 (f-1) Summary:
	mbedding. This relational word embedding is still learned from co-occurrence	knowledge graph embedding	word representations.	vector space word	Machine learning about	There is a large body of work on	Word embeddings have been	In @cite @cite, the authors
	tatistics, and can thus be used even when no external knowledge base is	model, to encourage the word	Machine learning about	representations are learned	language can be improved by	incorporating relational	widely used in NLP @cite @cite,	proposed an asymmetric cosine
	vailable. Our analysis shows that relational word vectors do indeed capture	vectors to additionally be	language can be improved by	from distributional information	supplying it with specific	knowledge into NLP tasks @cite	where they have been shown to	distance measure to emphasise
	formation that is complementary to what is encoded in standard word	predictive of a given set of	supplying it with specific	of words in large corpora.	knowledge and sources of	@cite. In particular, @cite	be useful for various NLP tasks,	the hypernymy relation of lexical
	mbeddings.	relational facts @cite @cite	knowledge and sources of	although such statistics are	external information. We	proposed a method to	such as word similarity, sentence	entailment. However, this
	2):	@cite . Other authors have used	external information. We	semantically informative, they disregard the valuable	present here a new version of	incorporate semantic knowledge	completion, name entity	distance measure does not
	espite the emergence and growth of numerous large knowledge graphs,	knowledge bases in a more restricted way, by taking the fact	present here a new version of the linked open data resource	information that is contained in	the linked open data resource ConceptNet that is particularly	into the learning process of word embeddings to improve	recognition, and so on @cite. However, these embeddings do	consider hypernymy directionality, hypernymy
	vailable on the Web. To address this, we present WebBrain, a new approach	that two words are linked to each	ConceptNet that is particularly	semantic lexicons such as	well suited to be used with	the quality of them. In @cite,	not explicitly capture relational	detection, or hypernymy
	or harvesting commonsense knowledge that relies on joint learning from	other in a given knowledge graph	well suited to be used with	WordNet, FrameNet, and the	modern NLP techniques such as	the authors proposed a method	information, which is the focus	classification. Furthermore, in
	Veb-scale data to fill gaps in the knowledge acquisition. We train a neural	as evidence that their word	modern NLP techniques such as	Paraphrase Database.	word embeddings. ConceptNet	for incorporating the	of this paper. In contrast, our	@cite the authors proposed a
ne	etwork model to learn relations based on large numbers of textual patterns	vectors should be similar @cite	word embeddings. ConceptNet	this paper proposes a novel	is a knowledge graph that	information about relationship	work aims to learn relational	method for using the cosine
	ound on the Web. At the same time, the model learns vector representations	@cite . Finally, there has also	is a knowledge graph that	framework for refining vector	connects words and phrases of	between entities that they	word embeddings that are	distance to emphasise
	f general word semantics. This joint approach allows us to generalize beyond	been work that uses lexicons to	connects words and phrases of	space representations using	natural language with labeled	extract from knowledge bases.	complementary to a given word	hypernymy, and they do not
	ne explicitly extracted information. Experiments show that we can obtain	learn word embeddings which	natural language with labeled	relational information from	edges. Its knowledge is collected	In this paper, we propose to	embedding, and can thus be used with modern NLP	consider the asymmetric distance measure. In @cite, the
	epresentations of words that reflect their semantics, yet also allow us to apture conceptual relationships and commonsense knowledge.	are specialized towards certain types of lexical knowledge, such	edges. Its knowledge is collected from many sources that include	semantic lexicons by encouraging linked words to	from many sources that include expert-created resources.	incorporate the information about relation between entities	techniques such as WordNet	authors presented a constrained
	RI:	as hypernymy @cite @cite .	expert-created resources.	have similar vector	crowd-sourcing, and games with	that is learned from	@cite and Word2Vec @cite to	ontimization problem, where the
	Ve present LEAR (Lexical Entailment Attract-Repel), a novel post-processing	antonymy @cite @cite or a	crowd-sourcing, and games with	representations, and it makes no	a purpose. It is designed to	distributional semantics alone.	learn relational embeddings.	data-derived objective function
	nethod that transforms any input word vector space to emphasise the	combination of various linguistic	a purpose. It is designed to	assumptions about how the	represent the general	Comment:	Comment:	is optimized subject to all ordinal
as	symmetric relation of lexical entailment (LE), also known as the IS-A or	constraints @cite .	represent the general	input vectors were constructed.	knowledge involved in	- Clear signs of MDS, with the 1st	- Clear signs of MDS, with the 1st	knowledge inequality constraints
	yponymy-hypernymy relation. By injecting external linguistic constraints (e.g.,		knowledge involved in	we present a novel post-	understanding language,	article (yellow) used for	article (yellow) used for	extracted from available
W	VordNet links) into the initial vector space, the LE specialisation procedure		understanding language,	processing method called LEAR	improving natural language	context (at the beginning),	context (at the beginning) and	knowledge resources such as
	rings true hyponymy-hypernymy pairs closer together in the transformed		improving natural language	(Lexical Entailment Attract-Repel ) that transforms any input word	applications by allowing the	with the 4th (purple) and 7th	its difference also clearly	Thesaurus and WordNet. Comment:
	uclidean space. The proposed asymmetric distance measure adjusts the orms of word vectors to reflect the actual WordNet-style hierarchy of		applications. This relational word embedding is still learned	vector space to emphasise the	application to better understand the meanings behind the words	(pink) articles are reference.  - The last sentence (red)	stated in the latter part.  - A part of the 4 <sup>th</sup> (purple) article	- Some signs of MDS, with the
	onepts. Simultaneously, a joint objective enforces semantic similarity using		from co-occurrence statistics.	asymmetric relation of lexical	people use. When ConceptNet is	attempts to contrast what is	is also leveraged to provide	3rd (blue) and 4th (purple)
	ne symmetric cosine distance, yielding a vector space specialised for both		and can thus be used even when	entailment (LE ), also known as	combined with word	being down in the 1st article	further reference.	articles are included in the
	exical relations at once. LEAR specialisation achieves state-of-the-art		no external knowledge base is	the hyponymy-hypernymy	embeddINGS acquired from	but used a phrase from the 6th	- While totally not accurate, it	summary.
	erformance in the tasks of hypernymy directionality, hypernymy detection,		available. Our analysis shows	relation. by injecting external	distributional semantics (such as	article (deep blue) instead	appears the model learned	- However, the red part is once
	nd graded lexical entailment, demonstrating the effectiveness and robustness		that relational word vectors do	linguistic constraints (e.g.,	word2vec), it provides		that WordNet and Word2Vec	again a hallucination
	f the proposed asymmetric specialisation model.		indeed capture information that	WordNet links ) into the initial	applications with understanding that they would not acquire		are entities related to word	
	1):  n this paper, we propose a general framework to incorporate semantic		is complementary to what is encoded in standard	vector space, the LE specialisation procedure brings	from distributional mechanics		embeddings and NLP techniques, thus copying them	
	nowledge into the popular data-driven learning process of word embeddings		words.    Despite the	true hyponymy - hypernymy	alone, nor from narrower		from the 6th article (deep blue)	
	improve the quality of them. Under this framework, we represent semantic		emergence and growth of	pairs closer together in the	resources such as WordNet or		nom the or drucke (deep blue)	
	nowledge as many ordinal ranking inequalities and formulate the learning of		numerous large knowledge	transformed Eucl	DBPedia. We demonstrate this			
	emantic word embeddings (SWE) as a constrained optimization problem,		graphs, many basic and	Comment:	with state-of-the-art results on			
	here the data-derived objective function is optimized subject to all ordinal		important facts about our	-Copied all but the last 2	intrinsic evaluations of word			
	nowledge inequality constraints extracted from available knowledge		everyday world are not readily	sentences of the 9th article,	relatedness that translate into			
	esources such as Thesaurus and WordNet. We have demonstrated that this onstrained optimization problem can be efficiently solved by the stochastic		available on the	then switched back to the 2nd (blue)	improvements on applications of word vectors, including solving			
	radient descent (SGD) algorithm, even for a large number of inequality		- All copied, but interestingly	- No indication of MDS	SAT-style analogies. In this paper,			
	constraints. Experimental results on four standard NLP tasks, including word		started form the last tokens of	- No indication of Wib3	we present a novel framework			
	milarity measure, sentence completion, name entity recognition, and the		the 5th article, copied the 6th,		called RC-NET to leverage both			
	OEFL synonym selection, have all demonstrated that the quality of learned		then switched back to the 2nd		the relational and categorical			
	rord vectors can be significantly improved after semantic knowledge is		- No indication of MDS		knowledge to produce word			
	ncorporated as inequality constraints during the learning process of word				representations of higher			
	mbeddings.				quality. Specifically, we build the			
	5): epresenting words into vectors in continuous space can form up a potentially				relational knowledge and Comment:			
	owerful basis to generate high-quality textual features for many text mining				- Copied the 6th article in full,			
	nd natural language processing tasks. Some recent efforts, such as the skip-				then switched to the middle oif			
gr	ram model, have attempted to learn word representations that can capture				the 5th (orange)			
	oth syntactic and semantic information among text corpus. However, they				- No indication of MDS			
	till lack the capability of encoding the properties of words and the complex							
	elationships among words very well, since text itself often contains							
	ncomplete and ambiguous information. Fortunately, knowledge graphs							
	rovide a golden mine for enhancing the quality of learned word epresentations. In particular, a knowledge graph, usually composed by							
	ntities (words, phrases, etc.), relations between entities, and some							
	orresponding meta information, can supply invaluable relational knowledge							
	nat encodes the relationship between entities as well as categorical							
	nowledge that encodes the attributes or properties of entities. Hence, in this							
pa	aper, we introduce a novel framework called RC-NET to leverage both the							

No.	Abstracts	Label	Base LED	Large LED	Base Centrum	Tuned LED	Tuned Centrum	Two-step
1872	(1):	The bad Nash equilibrium in	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:	Rouge scores:
	We consider the well-studied game-theoretic version of machine scheduling in	Example appears in several works	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:	- Rouge 2:
	which jobs correspond to self-interested users and machines correspond to	@cite @cite @cite to show	0.0592 (prec)	0.0329 (prec)	0.0884 (prec)	0.1442 (prec)	0.1259 (prec)	0.061 (prec)
	resources. Here each user chooses a machine trying to minimize her own cost,	that even for two machines the	0.0699 (recall)	0.035 (recall)	0.1329 (recall)	0.1049 (recall)	0.1259 (recall)	0.0699 (recall)
	and such selfish behavior typically results in some equilibrium which is not	price of anarchy is unbounded,	0.0641 (f-1)	0.0339 (f-1)	0.1061 (f-1)	0.1215 (f-1)	0.1259 (f-1)	0.0651 (f-1)
	globally optimal: An equilibrium is an allocation where no user can reduce her	thus suggesting that the notion	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:	- Rouge L:
	own cost by moving to another machine, which in general need not minimize the makespan, i.e., the maximum load over the machines. We provide tight	should be refined. Among these, the , which considers NE, is	0.1706 (prec) 0.2014 (recall)	0.183 (prec) 0.1944 (recall)	0.1389 (prec) 0.2083 (recall)	0.3143 (prec) 0.2292 (recall)	0.2153 (prec) 0.2153 (recall)	0.1758 (prec) 0.2014 (recall)
	bounds on two well-studied notions in algorithmic game theory, namely, the	studied in @cite @cite @cite .	0.1847 (f-1)	0.1944 (recail) 0.1886 (f-1)	0.1667 (f-1)	0.2651 (f-1)	0.2153 (f-1)	0.2014 (recall) 0.1877 (f-1)
	price of anarchy and the strong price of anarchy on machine scheduling setting	The . which considers equilibria	Summary:	Summary:	Summary:	Summary:	Summary:	Summary:
	which lies in between the related and the unrelated machine case. Both	arising in extensive form games,	We consider the well-studied	we study the outcome of	We consider the well-studied	There is a large body of	In @cite, Paes Leme, Syrgkanis,	In @cite, Paes Leme, Syrgkanis,
	notions study the social cost (makespan) of the worst equilibrium compared to	is studied in @cite @cite @cite	game-theoretic version of	natural learning algorithms in	game-theoretic version of	literature on the price of anarchy	and Tardos studied the price of	and Tardos studied games where
	the optimum, with the strong price of anarchy restricting to a stronger form of	@cite . In @cite the authors	machine scheduling in which	atomic congestion games.	machine scheduling in which	@cite @cite, which is closely	anarchy for atomic congestion	players choose their actions
	equilibria. Our results extend a prior study comparing the price of anarchy to	investigate equilibria and the	jobs correspond to self-	we consider the classical	jobs correspond to self-	related to our work. The price of	games. They showed that in	sequentially, and showed that
	the strong price of anarchy for two related machines (Epstein, Acta Informatica	resulting , while @cite focuses on	interested users and machines	machine scheduling game,	interested users and machines	anarchy is defined as the ratio of	almost all such games, the well-	sequential decisions mitigate the
	2010), thus providing further insights on the relation between these concepts.  Our exact bounds give a qualitative and quantitative comparison between the	the equilibria produced by the . A further distinction is between	correspond to resources. Here each user chooses a machine	where n jobs need to be scheduled on m machines, with	correspond to resources. Here each user chooses a machine	the worst strong Nash equilibria to the social optimum @cite. In	known multiplicative-weights learning algorithm results in	worst case outcomes known for the classical price of anarchy. In
	two models. The bounds also show that the setting is indeed easier than the	(randomized) and (deterministic)	trying to minimize her own cost,	the goal of minimizing the	trying to minimize her own cost,	particular, @cite and @cite	convergence to pure equilibria,	@cite, the authors consider a
	two unrelated machines: In the latter, the strong price of anarchy is @math ,	equilibria: in the former, players	and such selfish behavior	makespan, i.e. the maximum	and such selfish behavior	studied the trade-off between	and showed that if an authority	job scheduling game with two
	while in ours it is strictly smaller.	choose a probability distribution	typically results in some	load of any machine in the	typically results in some	the high and low price of	can change the order of the jobs	uniformly related parallel
	(2):	over the strategies and regard	equilibrium which is not globally	schedule.	equilibrium which is not globally	anarchy, and showed that the	adaptively to the decisions made	machines (or links). The goal of
	We study the outcome of natural learning algorithms in atomic congestion	their expected cost, in the latter	optimal: An equilibrium is an	the private goal of each job is to	optimal: An equilibrium is an	high price of anarchy can be	by the jobs so far (but cannot	each job is to minimize its own
	games. Atomic congestion games have a wide variety of equilibria often with	they choose deterministically one	allocation where no user can	minimize its own delay and the	allocation where no user can	bounded as a function of the	influence the decisions of the	delay and the social goal is to
	vastly differing social costs. We show that in almost all such games, the well-	strategy. In this work we focus on	reduce her own cost by moving	social goal is to minimize the	reduce her own cost by moving	number of machines and the	jobs), then there exists an	minimize the maximum delay of
	known multiplicative-weights learning algorithm results in convergence to pure equilibria. Our results show that nat- ural learning behavior can avoid bad	pure equilibria and in the remaining of this section we	to another machine, which in general need not minimize the	maximum delay of any job, that is, to minimize the makespan of	to another machine, which in general need not minimize the	size of the coalition. In the case of the network creation game	adaptive ordering in which the jobs end up in an optimum	any job, that is, to minimize the number of jobs that can be
	outcomes predicted by the price of anarchy in atomic congestion games such	write @math to denote the	makespan, i.e., the maximum	the machine running it.	makespan, i.e., the maximum	@cite the trade-offs between	schedule. They further extend	assigned to this machine. The
	as the load-balancing game introduced by Koutsoupias and Pa- padimitriou,	bounds on the price of anarchy	load over the machines. We	we show that in almost all such	load over the machines. We	the low and high prices of	their results to show that players	authors show that in almost all
	which has super-constant price of anarchy and has correlated equilibria that	for mixed equilibria.	provide tight bounds on two	games, the well- known	provide tight bounds on two	anarchy are studied.	can use algorithms with different	mixed Nash equilibria that are
	are exponentially worse than any mixed Nash equilibrium. Our results identify		well-studying notions in	multiplicative-weights learning	well-studies in algorithmic game	Comment:	(sufficiently small) learn-rate,	weakly stable, the Converse is
	a set of mixed Nash equilibria that we call weakly stable equilibria. Our notion		algorithmic game theory,	algorithm results in convergence	theory, namely, the price of	- Some signs of MDS, with the	i.e., they can trade off	true with probability 1 when
	of weakly stable is defined game-theoretically, but we show that this property		namely, the price of anarchy and	to pure equilibria.	anarchy and the strong price of	information in the 1st article	convergence speed and long	congestion costs are selected at
	holds whenever a stability criterion from the theory of dy- namical systems is		the strong price of anarchy on	furthermore, we show that if an	anarchy on machine scheduling	(yellow) used as context, and	term average regret differently.	random independently on each
	satisfied. This allows us to show that in every congestion game, the distribution of play converges to the set of weakly stable equilibria. Pure Nash		machine scheduling setting which lies in between the	authority can change the order of the jobs adaptively to the	setting which lies in between the related and the unrelated	the 3 <sup>rd</sup> (blue) and 4 <sup>th</sup> (purple) articles are included in the	In @cite @cite the authors studied the tradeoff between	edge (from any monotonically parametrized distribu-rontal)
	equilibria are weakly stable, and we show using techniques from algebraic		related and the unrelated	decisions made by the jobs so	machine case. Both notions	summary.	the quality of any Nash	The authors also extend their
	geometry that the converse is true with probability 1 when congestion costs		machine case. Both notions	far (but can not influence the	study the social cost (makespan)	,	equilibrium to the quality of a	results to the case of two
	are selected at random independently on each edge (from any monotonically		study the social cost (makespan)	decisions of the jobs ) then there	of the worst equilibrium		global optimum, and the	players, and show that for a
	parametrized distribu- tion). We further extend our results to show that		of the worst equilibrium	exists an adaptive ordering in	compared to the optimum, with		sequential price of anarchy.	wide range of speed ratios these
	players can use algorithms with different (sufficiently small) learn- ing rates,		compared to the optimum, with	which the jobs end up in an	the strong Price of anarchy		Comment:	two measures are very different
	i.e. they can trade off convergence speed and long term average regret		the strong price and the weak	optimum schedule.	restricting to a stronger form of		- Some signs of MDS, with the	whereas for other speed ratios
	differently.		price of anarchy restricting to a stronger form of equilibria. Our	Comment: - Very weak signs of MDS, with	equilibria. Our results extend a prior study comparing the price		2nd (green) ,4th (purple) and 5th (pink) articles summarized.	they give the exact same bound.  Comment:
	A strong equilibrium is a pure Nash equilibrium which is resilient to deviations		results extend a prior study	the 2nd (green), 4th (purple)	of Anarchy to the strong Price Of		This is some mixing of the ideas	- Some signs of MDS, with the
	by coalitions. We define the strong price of anarchy (SPoA) to be the ratio of		comparing the price of Anarchy	and 6th (orange) articles	Anarchy for two related		in the articles, possibly	2nd (green) ,3rd (green), 5th
	the worst strong equilibrium to the social optimum. Differently from the Price		to the strong price	covered.	machines (Epstein, Acta		because they used the same	(pink) and 6th (orange) articles
	of Anarchy (defined as the ratio of the worst Nash Equilibrium to the social		Comment:	-However, they are all	Informatica 2010), thus		special terms (e,.g. atomic	summarized.
	optimum), it quantifies the loss incurred from the lack of a central designer in		- Copied the first tokens only	presented as a single	providing further insights on the		congestion, price of anarchy)	- This is some mixing of the ideas
	settings that allow for coordination. We study the SPoA in two settings, namely		- No indication of MDS	document.	relation between these			in the articles, possibly
	job scheduling and network creation. In the job scheduling game we show that for unrelated machines the SPOA can be bounded as a function of the number			<ul> <li>The model also performs like an extractive model</li> </ul>	concepts. Our exact bounds give a qualitative and quantitative			because they used the same special terms (e,.g. atomic
	of machines and the size of the coalition. For the network creation game we			an extractive moder	comparison between the two			congestion, price of anarchy)
	show that the SPoA is at most 2. In both cases we show that a strong				models. The bounds also show			congestion, price of undienty
	equilibrium always exists, except for a well defined subset of network creation				that the setting is indeed easier			
	games.				than the two unrelated			
	(4):				machines: In the latter			
	We consider the classical machine scheduling, where n jobs need to be				Comment:			
	scheduled on m machines, with the goal of minimizing the makespan, i.e., the				Copied the first tokens only     No indication of MDS			
	maximum load of any machine in the schedule. We study inefficiency of schedules that are obtained when jobs arrive sequentially one by one, and				- NO IIIdication of MD3			
	choose themselves the machine on which they will be scheduled. Every job is							
	only interested to be on a machine with a small load (and does not care about							
	the loads of other machines). We measure the inefficiency of a schedule as the							
	ratio of the makespan obtained in the worst-case equilibrium schedule, and of							
	the optimum makespan. This ratio is known as the sequential price of anarchy.							
	We also introduce alternative inefficiency measures, which allow for a							
	favorable choice of the order in which the jobs make their decisions. We first disprove the conjecture of Hassin and Yovel (OR Letters, 2015) claiming that							
	for unrelated machines, i.e., for the setting where every job can have a							
	different processing time on every machine, the sequential price of anarchy for							
	m = 2 machines is at most 3. We provide an answer for the setting with m = 2							
ı	and show that the sequential price of anarchy grows at least linearly with the							
ı	number of players. Furthermore, we show that for a certain order of the jobs,							
1	the resulting makespan is at most linearly larger than the optimum makespan.							
	Furthermore, we show that if an authority can change the order of the jobs	l						