Qualitative Analysis of Generated Summaries

Introduction

1. This document provides a qualitative analysis of the summaries written by 6 models tested in the W266 project, namely:

   a) Baseline LED model ("allenai/led-base-16384"), with max input token length of 16384 & ~200M parameters
   b) Off-the-shelf LED model ("allenai/led-large-16384-arxiv")[1], with max input token length of 16384 & ~512M parameters
   c) Off-the-shelf LED Centrum model ("ratishsp/Centrum"), with max input token length of 4096[2] & ~192M parameters
   d) Finetuned LED model (our own model), with max input token length of 16384 but finetuned using inputs tokenized at 4096 max
   e) Finetuned Centrum model (our own model), with max input token length of 4096
   f) Two-step LED-to-Centrum model (our own model), with max input token length of 4096 at the second step[3]

2. A total of 25 randomly drawn samples[4] (from the 5093 samples in the X-Science test dataset) are analyzed. These 25 samples include:

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

3. In analyzing the results, a few dimensions are looked at:

   a) general fluency, i.e. whether the summary looks coherent and contains blatant errors or strange symbols
   b) indication of multi-document summary ("MDS"), e.g. multiple sentences each attempting to summarize the work done in one or more of the journals being referenced
   c) validity of MDS, e.g. whether sentences appearing to be summarizing different journals are in fact extracting key information from that journal's abstract
   d) in addition, highest scores are highlighted in green, while lowest ones are in red, while key ideas from different articles are highlighted in different colors

High-level summary

4. As set out in the main paper, this analysis provides us with further insights, and we note the following:

   a) in general, the fine-tuned LED, fine-tuned Centrum and two-step model manages to summarize from multiple sources most of the time (see the different colors for excerpts for different parts of the source)
   b) the writings from these three models are usually very fluent as well, though occasional strange formulations and factual inaccuracies are observed
   c) furthermore, these models show clear signs of learning the desired writing style, especially in contrasting the main article with the other references
   d) as a contrast, the baseline LED is prone to just copying from just the first part of the main article; while the off-the-shelf LED and Centrum either just copies or sometimes extracts from multiple sources

Results *[only those referred to in the main paper included; full version of this document with all 25 samples available at Github.]*

(a) Short samples

- 3157, 4820: Samples with used to illustrate the presence of long and compound nouns

- 4160: Short example where the off-the-shelf LED and off-the-shelf Centrum managed to extract information from multiple sources as well (i.e. in addition to the tuned and two-step models)

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
3157	**(1):** In this paper, we study a novel approach for named entity recognition (NER) and mention detection in natural language processing. Instead of treating NER as a sequence labelling problem, we propose a new local detection approach, which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left right contexts into a fixed-size representation. Afterwards, a simple feedforward neural network is used to reject or predict entity label for each individual fragment. The proposed method has been evaluated in several popular NER and mention detection tasks, including the CoNLL 2003 NER task and TAC-KBP2015 and TAC-KBP2016 Tri-lingual Entity Discovery and Linking (EDL) tasks. Our methods have yielded pretty strong performance in all of these examined tasks. This local detection approach has shown many advantages over the traditional sequence labelling methods. **(2):**	The success of word embedding @cite encourages researchers to focus on machine-learned representation instead of heavy feature engineering in NLP. Using word embedding as the typical feature representation for words, NNs become competitive to traditional approaches in NER. Many NLP tasks, such as NER, chunking and part-of-speech (POS) tagging can be formulated as sequence labeling tasks. In @cite , deep convolutional neural networks (CNN) and conditional random fields (CR	**Rouge scores:** - Rouge 2: 0.0316 (prec) 0.0463 (recall) 0.0376 (f-1) - Rouge L: 0.1132 (prec) 0.1651 (recall) 0.1343 (f-1) **Summary:** In this paper, we study a novel approach for named entity recognition (NER) and mention detection in natural language processing. Instead of treating NER as	**Rouge scores:** - Rouge 2: 0.0303 (prec) 0.037 (recall) 0.0333 (f-1) - Rouge L: 0.0977 (prec) 0.1193 (recall) 0.1074 (f-1) **Summary:** in this paper, we study a novel approach for named entity recognition and mention detection in natural language processing. instead of treating named entity	**Rouge scores:** - Rouge 2: 0.035 (prec) 0.0648 (recall) 0.0455 (f-1) - Rouge L: 0.1045 (prec) 0.1927 (recall) 0.1355 (f-1) **Summary:** In this paper, we study a novel approach for named entity recognition (NER) and mention detection in natural language processing. Instead of treating NER as a	**Rouge scores:** - Rouge 2: 0.0685 (prec) 0.0463 (recall) 0.0552 (f-1) - Rouge L: 0.1757 (prec) 0.1193 (recall) 0.1421 (f-1) **Summary:** In @cite, the authors propose a unified neural network architecture and learning algorithm that can be applied to various natural language processing tasks incl	**Rouge scores:** - Rouge 2: 0.0 (prec) 0.0 (recall) 0.0 (f-1) - Rouge L: 0.1957 (prec) 0.0826 (recall) 0.1161 (f-1) **Summary:** The Skip-gram model @cite @cite is an efficient method for learning high-quality distributed vector representations that capture a large number of precise synt	**Rouge scores:** - Rouge 2: 0.049 (prec) 0.0463 (recall) 0.0476 (f-1) - Rouge L: 0.1942 (prec) 0.1835 (recall) 0.1887 (f-1) **Summary:** In @cite, the authors proposed a unified neural network architecture and learning algorithm that can be applied to various natural language processing tasks i

[1] This model has been pre-trained by the authors for general summarization task using arXiv dataset.
[2] 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.
[3] 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 finetuned LED) receiving truncated inputs.
[4] For details, please refer to the "Qualitative_analysis.ipynb" notebook. We originally planned for 20 samples for each category but ended up doing 25 in total due to time limitations. The 10 are the first and last 5 of the randomly selected samples for the short and medium samples, and just the first 5 for the long samples.

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
	We propose a unified neural network architecture and learning algorithm that can be applied to various natural language processing tasks including: part-of-speech tagging, chunking, named entity recognition, and 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 features carefully optimized for each task, our system learns internal representations on the basis of vast amounts of mostly unlabeled training data. This work is then used as a basis for building a freely available tagging system with good performance and minimal computational requirements. **(3):** The recently introduced continuous Skip-gram model is an efficient method for learning high-quality distributed vector representations that capture a large number of precise syntactic and semantic word relationships. In this paper we present several extensions that improve both the quality of the vectors and the training speed. By subsampling of the frequent words we obtain significant speedup and also learn more regular word representations. We also describe a simple alternative to the hierarchical softmax called negative sampling. An inherent limitation of word representations is their indifference to word order and their inability to represent idiomatic phrases. For example, the meanings of "Canada" and "Air" cannot be easily combined to obtain "Air Canada". Motivated by this example, we present a simple method for finding phrases in text, and show that learning good vector representations for millions of phrases is possible.	F) are used to infer NER labels at a sentence level, where they still use many hand-crafted features to improve performance, such as capitalization features explicitly defined based on first-letter capital, non-initial capital and so on.	a sequence labelling problem, we propose a new local detection approach, which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left right contexts into a fixed-size representation. Afterwards, a simple feedforward neural network is used to reject or predict entity label for each individual fragment. The proposed method has been evaluated in several popular NER and mention detection tasks, including the CoNLL 2003 NER task and TAC-KBP2016 Tri-lingual Entity Discovery and Linking (EDL) tasks. Our methods have yielded pretty strong performance in all of these examined tasks. This local detection approach has shown many advantages over the traditional sequence labelling methods.|||||We propose a unified neural network architecture and learning algorithm that can be applied to various natural language processing **Comment:** - Copied the first tokens only - No indication of MDS	recognition as a sequence labelling problem, we propose a new local detection approach, which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left right contexts into a fixed-size representation. afterwards, a simple feedforward neural network is used to reject or predict entity label for each individual fragment. the proposed method has been evaluated in several popular named entity and mention detection tasks, including the coNLL 2003 named entity recognition task and the tac-2015 and tac-2016 tri-lingual entity discovery and linking tasks. our methods have yielded pretty strong performance in all of these examined tasks. **Comment:** - Copied the first tokens only, except for the deletion of the acronym for NER (red) - No indication of MDS	sequence labelling problem, we propose a new local detection approach, which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encoding each sentence fragment and its left right contexts into a fixed-size representation. Afterwards, a simple feedforward neural network is used to reject or predict entity label for each individual fragment. The proposed method has been evaluated in several popular NER and mention detection tasks, including the CoNLL 2003 NER task and TAC-KBP2015 and TAC -KBP2016 Tri-lingual Entity Discovery and Linking (EDL) tasks. Our methods have yielded pretty strong performance in all of these examined tasks. This local detection approach has shown many advantages over the traditional sequence labelling methods. We propose a unified neural network architecture and learning algorithm that can be applied to various natural language processing tasks including: part-of-speech tagging, chunking, named entity recognition, and 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:** - Copied the first tokens only - No indication of MDS	uding: part-of-speech tagging, chunking, named entity recognition, and semantic role labeling. In this work, we propose a new local detection approach, which rely on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left right contexts into a fixed-size representation. **Comment:** - Shows signs of MDS, which contrasts the 2nd (green) article with the 1st (yellow) one. - However, the 3rd article is not covered at all	actic and semantic word relationships. By subsampling of the frequent words, it obtain significant speedup and also learn more regular word representations. **Comment:** - Shows weak signs of MDS, providing a summary of the 3rd article only (blue) and presenting it as different from the main one.	ncluding part-of-speech tagging, chunking, named entity recognition, and semantic role labeling. However, unlike our work, they do not use a deep neural network to train their system. Instead, they use a multi-task learning algorithm to train their network on a large amount of unlabeled training data. In contrast, we use a deep network to train our system on a much larger amount of training data, and we do not need to train our network on unlabeled data. **Comment:** - Shows signs of MDS, providing a summary of the 2nd article (green) and contrasting it with the main one (red). - However, the red parts also contain hallucinations, not to mention the neglect of the 3rd article
4160	**(1):** 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 cross-based cost aggregation and semiglobal matching, followed by a left-right consistency check to eliminate errors in the occluded 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. We present a supervised learning approach for predicting the correctness of stereo matches based on a random forest and a set of features that capture various forms of information about each pixel. We show highly competitive results in predicting the correctness of matches and in confidence estimation, which allows us to rank pixels 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 an important distinction from current literature that has mainly focused on sparsification by removing potentially erroneous disparities to generate quasi-dense disparity maps. **(3):** With the aim to improve accuracy of stereo confidence measures, we apply the random decision forest framework to a large set of diverse stereo confidence measures. Learning and testing sets were drawn from the recently introduced KITTI dataset, which currently poses higher challenges to stereo solvers than other benchmarks with ground truth for stereo evaluation. We experiment with semi global matching stereo (SGM) and a census data term, which is the best performing real-time capable stereo method known to date. On KITTI images, SGM still produces a significant amount of error. We obtain consistently improved area under curve values of sparsification measures in comparison to best performing single stereo confidence measures where numbers of stereo errors are large. More specifically, our method performs best in all but one out of 194 frames of the KITTI dataset.	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: 0.0523 (prec) 0.1579 (recall) 0.0786 (f-1) - Rouge L: 0.0983 (prec) 0.2931 (recall) 0.1472 (f-1) **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 cross-based cost aggregation and semiglobal matching, followed by a left-right consistency check to eliminate errors in the occluded 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.|||||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. We present a supervised learning approach for pre	**Rouge scores:** - Rouge 2: 0.0826 (prec) 0.1579 (recall) 0.1084 (f-1) - Rouge L: 0.2 (prec) 0.3793 (recall) 0.2619 (f-1) **Summary:** we present a supervised learning approach for predicting the correctness of stereo matches based on a random forest and a set of features that capture various forms of information about each pixel. 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 cross-based cost aggregation and semiglobal matching, followed by a left-right consistency check to eliminate errors in the occluded regions. our stereo method achieves an error rate of 2.61 on the recently introduced kitti stereo dataset and is currently (August 2014) the top performing method on this dataset. **Comment:**	**Rouge scores:** - Rouge 2: 0.0468 (prec) 0.1404 (recall) 0.0702 (f-1) - Rouge L: 0.1047 (prec) 0.3103 (recall) 0.1565 (f-1) **Summary:** In this article, we're going to take a look at a new method for predicting the correctness of stereo matches based on a random forest and a set of features that capture various forms of information about each pixel. We're using the random decision forest framework, which currently poses higher challenges to stereo solvers than other benchmarks with ground truth for stereo evaluation. We experiment with semi global matching stereo (SGM) and a census data term, which is the best performing real-time capable stereo method known to date. On KITTI images, SGM still produces a significant amount of error. We obtain consistently improved area under curve values of sparsification measures in comparison to best performing single stereo confidence measures where numbers of ste	**Rouge scores:** - Rouge 2: 0.1515 (prec) 0.1754 (recall) 0.1626 (f-1) - Rouge L: 0.2687 (prec) 0.3103 (recall) 0.288 (f-1) **Summary:** Our work is closely related to @cite @cite in that they use a random forest to predict the correctness of stereo matches based on a set of features that capture various forms of information about each pixel. However, our work is different in that we train a convolutional neural network to predict how well two image patches match and use it to compute the stereo matching cost. **Comment:** - Clear sign of MDS, with the sentence in red summarizing both the 2nd and 3rd articles (both uses random forests) and contrasting that with a fact from the 1st (yellow)	**Rouge scores:** - Rouge 2: 0.0517 (prec) 0.0526 (recall) 0.0522 (f-1) - Rouge L: 0.2373 (prec) 0.2414 (recall) 0.2393 (f-1) **Summary:** Our work is also related to the semi-global matching stereo (SGM) method @cite @cite, which is based on a random forest and a set of features that capture various forms of information about each pixel. SGM achieves an error rate of 2.61 on the KITTI stereo dataset and is currently the top performing method on this dataset. **Comment:** - Some signs of MDS, with the first green part showing an attempt to make comparison - The blue and 2nd green part shows possible mix up between the 2nd and 3rd models, caused probably by the repeated use of stereo and SGM - In a similar vein, the red part is from the 1st article (see numbers in yellow) but the sentence structure is unclear	**Rouge scores:** - Rouge 2: 0.1122 (prec) 0.193 (recall) 0.1419 (f-1) - Rouge L: 0.2121 (prec) 0.3621 (recall) 0.2675 (f-1) **Summary:** In @cite @cite, the authors used a random forest to predict the correctness of stereo matching based on a random forest and a set of features that capture various forms of information about each pixel. However, they do not consider sparsification. In addition, they did not use sparsification to improve the accuracy of the disparity maps, and they do not require sparsification in order to improve the performance of the matching. Moreover, their work does not rely on the sparsification of sparsification, but rather on sparsifying the disparity maps in order to achieve sparsification without sparsification and without sparsifying. **Comment:** - Clear sign of MDS, with the first sentence referring to the 2nd and 3rd examples, and

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
			dicting the correctness of stereo matches based on a random forest and a set of features that capture various forms of information about each pixel. We show highly competitive results in predicting the accuracy of matches and in confidence estimation, which allows us to rank **Comment:** - Copied the first tokens only - No indication of MDS	- Copied from the 1st (yellow) and 2nd (green) articles - No indication of MDS	reo errors are large. More specifically, our method performs best in all but one out of 194 frames of the KITTI dataset. This is an important distinction from current literature that has mainly focused on sparsification by removing potentially erroneous disparities to generate quasi-dense disparity maps. **Comment:** - Some sign of MDS by contrasting the 2nd (green) and 3rd (blue articles) - However, the phrase in red is also copied from the 2nd article so it is uncertain if the model is actually doing MDS - It also mixed up the facts in the 2nd and 3rd articles		as to whether the model is referring to the main article or not.	contrasting it with the main one. - However, the red parts are hallucinations
4820	**(1):** This paper presents a new method and a constraint-based objective function to solve two problemsrelated to the design of optical telecommunication networks, namely the Synchronous Optical Net-work Ring Assignment Problem (SRAP) and the Intra-ring Synchronous Optical Network DesignProblem (IDP). These network topology problems can be represented as a graph partitioning withcapacity constraints as shown in previous works. W`e present here a new objective function and anew local search algorithm to solve these problems. Experiments conducted in C **(2):** This paper considers two problems that arise in the design of optical telecommunication networks when a ring-based topology is adopted, namely the SONET Ring Assignment Problem and the Intraring Synchronous Optical Network Design Problem. We show that these two network topology problems correspond to graph partitioning problems with capacity constraints: the first is a vertex partitioning problem, while the latter is an edge partitioning problem. We consider solution methods for both problems, based on metaheuristic algorithms. We first describe variable objective functions that depend on the transition from one solution to a neighboring one, then we apply several diversification and intensification techniques including Path Relinking, eXploring Tabu Search and Scatter Search. Finally we propose a diversification method based on the use of multiple neighborhoods. A set of extensive computational results is used to compare the behaviour of the proposed methods and objective functions.	The two other methods described in @cite are based on the idea that to save ADMs a good solution should have very dense rings. They are both greedy and rely on a clique algorithm. In graph theory, a clique in an undirected graph @math is a subset of the vertex set @math , such that for every two vertices in @math , there exists an edge connecting the two. Finding a clique is not that easy, a way to do it is to use an "Union-Find" strategie, two clique @math and @math such that each node in @math is adjacent to each node in @math then merge the two cliques (). The associated heuristic starts by considering each node to be a clique of size one, and to merge two cliques into a larger clique until there are no more possible merges.	**Rouge scores:** - Rouge 2: 0.0132 (prec) 0.0144 (recall) 0.0138 (f-1) - Rouge L: 0.1316 (prec) 0.1429 (recall) 0.137 (f-1) **Summary:** This paper presents a new method and a constraint-based objective function to solve two problemsrelated to the design of optical telecommunication networks, namely the Synchronous Optical Net-work Ring Assignment Problem (SRAP) and the Intra-ring Synchronous optical Network DesignProblem (IDP). These network topology problems can be represented as a graph partitioning withcapacity constraints as shown in previous works. We present here a new objective function and anew local search algorithm to solve these problems. Experiments conducted in CERN and the University of California, Berkeley.|This paper considers two problems that arise in the design of Optical telecommunication networks when a ring-based topology is adopted, namely the SONET Ring Assignment Problem and the Intraring Synchronous O Network Design Problem. We show that these two network topologies problems correspond to graph partitioning problems with capacity constraints: the first is a vertex partitioning problem, while the latter is an edge partitioning problem. **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0 (prec) 0.0 (recall) 0.0 (f-1) - Rouge L: 0.1579 (prec) 0.0857 (recall) 0.1111 (f-1) **Summary:** this paper presents a new method and a constraint-based objective function to solve two problemsrelated to the design of optical telecommunication networks, namely the Synchronous Optical Net-work Ring Assignment Problem (SRAP) and the Intra-ring Synchronous optical Network DesignProblem (IDp). these network topology problems can be represented as a graph partitioning withcapacity constraints as shown in previous works. we present here a new objective function and anew local search algorithm to solve these problems. **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0205 (prec) 0.0288 (recall) 0.024 (f-1) - Rouge L: 0.1276 (prec) 0.1786 (recall) 0.1488 (f-1) **Summary:** This paper presents a new method and a constraint-based objective function to solve two problems related to the design of optical telecommunication networks, namely the Synchronous Optical Net-work Ring Assignment Problem (SRAP) and the Intra-ring Synchronous optical Network DesignProblem (IDP). These network topology problems can be represented as a graph partitioning withcapacity constraints as shown in previous works. We present here a new objective function and anew local search algorithm to solve these problems. Experiments conducted in C This paper considers two problems that arise in the design of Optical telecommunication networks when a ring-based topology is adopted, namely the SONET Ring Assignment Problem and the Intra-ring SynchronousOptical Network Design Problem. We show that these two network topology problem correspond to graph partitioning with capacity constraints: the first is a partitioning problem, while the latter is an edge partitioning problem. We consider solution methods for both problems, based on metaheuristic algorithms. We first describe variable objective functions that depend on the transition from one solution to a neighboring one, then we apply several diversification and intensification techniques including Path Relinking, eXploring Tabu Search and Scatter Search. Finally we propose a divers **Comment:** - Copied the first tokens only	**Rouge scores:** - Rouge 2: 0.0189 (prec) 0.0072 (recall) 0.0104 (f-1) - Rouge L: 0.2407 (prec) 0.0929 (recall) 0.134 (f-1) **Summary:** In the context of optical telecommunication networks, the SONET Ring Assignment Problem (SRAP) and the Intraring Synchronous Optical Network Design Problem (IDP) are two problems that arise when a ring-based topology is adopted. The SRAP problem can be represented as a graph partitioning problem, while the IDP problem is a edge partitioning problem. **Comment:** - Shows no sign of MDS, with the model just extracting the common theme of the 2 articles	**Rouge scores:** - Rouge 2: 0.0357 (prec) 0.0072 (recall) 0.012 (f-1) - Rouge L: 0.2414 (prec) 0.05 (recall) 0.0828 (f-1) **Summary:** In @cite, an objective function and a diversification method are proposed to solve the SONET Ring Assignment Problem (SRAP) and the Intra-ring Synchronous Optical Network Design Problem (IDP). **Comment:** - Shows weak signs of MDS, which presents the 2nd (green) article as different from the main one	**Rouge scores:** - Rouge 2: 0.1233 (prec) 0.0647 (recall) 0.0849 (f-1) - Rouge L: 0.2703 (prec) 0.1429 (recall) 0.1869 (f-1) **Summary:** In @cite, the objective functions of the SONET ring assignment problem and the IDP problem are discussed in terms of a metaheuristic algorithm. In particular, the objective function for SONET Ring Assignment Problem is described in @cite. The objective functions of SONet Ring Assignment Problem are defined as: where @math is the number of neighbors, @math the number of nodes in the network, @math and @math are the number of vertices in the network. **Comment:** - Shows weak signs of MDS, which presents the 2nd (green) article as different from the main one - Contains more relevant information that the other tuned models

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
					- No indication of MDS			

(b) Medium samples

- 831, 5068: examples where the tuned and two-step models show the ability to contrast and compare the main article with other references.

- 845: sample where the tuned models avoided quoting from the irrelevant input

- 4371: sample where even the off-the-shelf Centrum resorted to just copying from one of the articles

- 4858: sample where highly technical language use is present causing the models to show weaker MDS performance

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

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
845	**(1):** We present a neural framework for opinion summarization from online product reviews which is knowledge-lean and only requires light supervision (e.g., in the form of product domain labels and user-provided ratings). Our method combines two weakly supervised components to identify salient opinions and form extractive summaries from multiple reviews: an aspect extractor trained under a multi-task objective, and a sentiment predictor based on multiple instance learning. We introduce an opinion summarization dataset that includes a training set of product reviews from six diverse domains and human-annotated development and test sets with gold standard aspect annotations, salience labels, and opinion summaries. Automatic evaluation shows significant improvements over baselines, and a large-scale study indicates that our opinion summaries are preferred by human judges according to multiple criteria. **(2):** The specification discloses a luggage carrier made up of a generally U-shaped frame. The frame has two spaced legs with a hook on the front which hooks over the bumper of an automobile. Two braces are attached to the cross member of the U-shaped member and the front portion of the braces is received on fastening means welded to the under side of the car frame. The cross members provide a supporting surface for carrying articles, boats and the like. A platform may be supported on the frame. **(3):** We develop a Ranking framework upon Recursive Neural Networks (R2N2) to rank sentences for multi-document summarization. It formulates the sentence ranking task as a hierarchical regression process, which simultaneously measures the salience of a sentence and its constituents (e.g., phrases) in the parsing tree. This enables us to draw on word-level to sentence-level supervisions derived from reference summaries. In addition, recursive neural networks are used to automatically learn ranking features over the tree, with hand-crafted feature vectors of words as inputs. Hierarchical regressions are then conducted with learned features concatenating raw features. Ranking scores of sentences and words are utilized to effectively select informative and non-redundant sentences to generate summaries. Experiments on the DUC 2001, 2002 and 2004 multi-document summarization datasets show that R2N2 outperforms state-of-the-art extractive summarization approaches. **(4):** We propose a neural multi-document summarization (MDS) system that incorporates sentence relation graphs. We employ a Graph Convolutional Network (GCN) on the relation graphs, with sentence embeddings obtained from Recurrent Neural Networks as input node features. Through multiple layer-wise propagation, the GCN generates high-level hidden sentence features for salience estimation. We then use a greedy heuristic to extract salient sentences while avoiding redundancy. In our experiments on DUC 2004, we consider three types of sentence relation graphs and demonstrate the advantage of combining sentence relations in graphs with the representation power of deep neural networks. Our model improves upon traditional graph-based extractive approaches and the vanilla GRU sequence model with no graph, and it achieves competitive results against other state-of-the-art multi-document summarization systems.	A few extractive neural models have been recently applied to generic multi-document summarization. train a recursive neural network using a ranking objective to identify salient sentences, while follow-up work @cite employs a multi-task objective to improve sentence extraction, an idea we adapted to our task. propose a graph convolution network to represent sentence relations and estimate sentence salience. Our summarization method is tailored to the opinion extraction task, it identifies aspect-specific and salient units, while minimizing the redundancy of the final summary with a greedy selection algorithm @cite @cite . Redundancy is also addressed in who propose a graph-based framework for abstractive summarization. introduce an encoder-decoder neural method for extractive opinion summarization. Their approach requires direct supervision via gold-standard extractive summaries for training, in contrast to our weakly supervised formulation.	**Rouge scores:** - Rouge 2: 0.0732 (prec) 0.0889 (recall) 0.0803 (f-1) - Rouge L: 0.1394 (prec) 0.1691 (recall) 0.1528 (f-1) **Summary:** We present a neural framework for opinion summarization from online product reviews which is knowledge-lean and only requires light supervision (e.g., in the form of product domain labels and user-provided ratings). Our method combines two weakly supervised components to identify salient opinions and form extractive summaries from multiple reviews: an aspect extractor trained under a multi-task objective, and a sentiment predictor based on multiple instance learning. We introduce an opinion summarization dataset that includes a training set of product reviews from six diverse domains and human-annotated development and test sets with gold standard aspect annotations, salience labels, and opinion summaries. Automatic evaluation shows significant improvements over baselines, and a large-scale study indicates that our opinion summaries are preferred by human judges according to multiple criteria.|||||The specification discloses a luggage carrier made up of a generally U-shaped frame. The frame has two spaced legs with a hook on the front which hooks over the bumper of an automobile **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0902 (prec) 0.0815 (recall) 0.0856 (f-1) - Rouge L: 0.1707 (prec) 0.1544 (recall) 0.1622 (f-1) **Summary:** we present a neural framework for opinion summarization from online product reviews which is knowledge-lean and only requires light supervision (e.g., in the form of product domain labels and user-provided ratings). our method combines two weakly supervised components to identify salient opinions and form extractive summaries from multiple reviews : an aspect extractor trained under a multi-task objective, and a sentiment predictor based on multiple instance learning. we introduce an opinion summarization dataset that includes a training set of product reviews from six diverse domains and human-annotated development and test sets with gold standard aspect annotations, salience labels, and opinion summaries. a large-scale study indicates that our opinion summaries are preferred by human judges according to multiple criteria. **Comment:** - Copied the first tokens only, except that the phrase "Automatic evaluation shows significant improvements over baselines, and" is skipped before the red part - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0651 (prec) 0.1037 (recall) 0.08 (f-1) - Rouge L: 0.1157 (prec) 0.1838 (recall) 0.142 (f-1) **Summary:** We present a neural framework for opinion summarization from online product reviews which is knowledge-lean and only requires light supervision (e.g., in the form of product domain labels and user-provided ratings). Our method combines two weakly supervised components to identify salient opinions and form extractive summaries from multiple reviews: an aspect extractor trained under a multi-task objective, and a sentiment predictor based on multiple instance learning. We introduce an opinion summarization dataset that includes a training set of product reviews from six diverse domains and human-annotated development and test sets with gold standard aspect annotations, salience labels, and opinion summaries. Automatic evaluation shows significant improvements over baselines, and a large-scale study indicates that our opinion summaries are preferred by human judges according to multiple criteria. The specification discloses a luggage carrier made up of a generally U-shaped frame. The frame has two spaced legs with a hook on the front which hooks over the bumper of an automobile. Two braces are attached to the cross member of the U-shaped member and the front portion of the braces is received on fastening means welded to the under side of the car frame. The cross members provide a supporting surface for carrying articles, boats and the like. A platform may **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.1695 (prec) 0.0741 (recall) 0.1031 (f-1) - Rouge L: 0.3333 (prec) 0.1471 (recall) 0.2041 (f-1) **Summary:** In the context of multi-document summarization, R2N2 @cite is a recurrent neural network (RNN) for extracting salience labels. The RNN is trained on sentence relation graphs, and the salience labels are extracted using a greedy heuristic to extract salient sentences while avoiding redundancy @cite @cite. In this work, we use the RNN to extract salience labels from product reviews. **Comment:** - Shows clear sign of MDS, summarizing the 3rd (blue) and 4th (purple) articles, while clearly contrasting the main article (yellow). - The RNN notation proved slightly confusing for the model though, the 2nd article used "Recursive Neural Network", not "Recurrent Neural Network" - The 2nd article appears completely unrelated to the task, and the model probably correctly omits it.	**Rouge scores:** - Rouge 2: 0.1186 (prec) 0.0519 (recall) 0.0722 (f-1) - Rouge L: 0.2167 (prec) 0.0956 (recall) 0.1327 (f-1) **Summary:** There is a large body of work on multi-document summarization, including @cite @cite, @cite and @cite. However, most of these works focus on extracting features from a single review. In this work, we focus on extractive summaries from multiple reviews: an aspect extractor trained under a multi-task objective, and a sentiment predictor trained based on multiple instance learning. **Comment:** - Shows some sign of MDS, with the model summarizing what the main article does (yellow), and trying to contrast that with other literature (red) - However, the red part is actually not factually based, and the phrase "a large body of work" is probably picked up from the summarization training.	**Rouge scores:** - Rouge 2: 0.1594 (prec) 0.0815 (recall) 0.1078 (f-1) - Rouge L: 0.2429 (prec) 0.125 (recall) 0.165 (f-1) **Summary:** In the context of summarization, @cite employ a Graph Convolutional Network (GCN) on the relation graphs, with sentence embeddings from Recurrent Neural Networks as input node features for salience estimation. However, they do not consider sentence relation graphs and do not consider the representation power of deep neural networks. In contrast, we use salience labels and salience labels to identify salient opinions and form extractive summaries from multiple reviews. **Comment:** - Shows some sign of MDS, the model summarizing what the main article does (yellow), and trying to contrast that with other literature (purple) - However, the red-and-purple part contains erroneous facts. The model wrongly assumes that the 4th article did not look at sentence relation graphs when the authors explicitly say they did.
4371	**(1):** In contrast to XML query languages as e.g. XPath which require knowledge on the query language as well as on the document structure, keyword search is open to anybody. As the size of XML sources grows rapidly, the need for efficient search indices on XML data that support keyword search increases. In this paper, we present an approach of XML keyword search which is based on the DAG of the XML data, where repeated substructures are considered only once, and therefore, have to be searched only once. As our performance evaluation shows, this DAG-based extension of the set intersection search algorithm [1], [2], can lead to search times that are on large documents more than twice as fast as the search times of the XML-based approach. Additionally, we utilize a smaller index, i.e., we consume less main memory to compute the results. **(2):** Recent works have shown the benefits of keyword proximity search in querying XML documents in addition to text documents. For example, given query k	Similar to XRANK @cite is the stack-based approach presented in @cite . In contrast to the previous approaches, the authors do not used the DeweyID to identify a node and to calculate the ancestor-descendant or even parent-child relationships, but they propose to use a combination of preorder position, postorder position, and depth of the node.	**Rouge scores:** - Rouge 2: 0.0244 (prec) 0.069 (recall) 0.036 (f-1) - Rouge L: 0.0788 (prec) 0.2203 (recall) 0.1161 (f-1) **Summary:** In contrast to XML query languages as e.g. XPath which require knowledge on the query language as well as on the document structure, keyword search is ope	**Rouge scores:** - Rouge 2: 0.0 (prec) 0.0 (recall) 0.0 (f-1) - Rouge L: 0.1282 (prec) 0.1695 (recall) 0.146 (f-1) **Summary:** in this paper, we consider the problem of efficiently producing ranked results for keyword search queries over hyperlinked XML documents.	**Rouge scores:** - Rouge 2: 0.0061 (prec) 0.0172 (recall) 0.009 (f-1) - Rouge L: 0.0793 (prec) 0.2203 (recall) 0.1166 (f-1) **Summary:** "We consider the problem of efficiently producing ranked results for keyword search queries over hyperlinked Xie documents, as opposed to (conceptually) fla	**Rouge scores:** - Rouge 2: 0.0429 (prec) 0.0517 (recall) 0.0469 (f-1) - Rouge L: 0.1831 (prec) 0.2203 (recall) 0.2 (f-1) **Summary:** XRANK @cite is an extension of the set intersection search algorithm @cite, which is based on the DAG of the XML data. XRANK is a hyperlink based HTML sear	**Rouge scores:** - Rouge 2: 0.125 (prec) 0.0517 (recall) 0.0732 (f-1) - Rouge L: 0.4 (prec) 0.1695 (recall) 0.2381 (f-1) **Summary:** XRANK @cite is an extension of the set intersection search algorithm @cite. XRANK can be used to query a mix of HTML and XML documents.	**Rouge scores:** - Rouge 2: 0.04 (prec) 0.0862 (recall) 0.0546 (f-1) - Rouge L: 0.1349 (prec) 0.2881 (recall) 0.1838 (f-1) **Summary:** XRANK @cite is a hyperlink based HTML search engine. It is designed to perform keyword search over hyperlinked XML documents. However, it does not consi

<table>
<tr><th>No.</th><th>Abstracts</th><th>Label</th><th>Base LED</th><th>Off-the-shelf LED</th><th>Off-the-shelf Centrum</th><th>Tuned LED</th><th>Tuned Centrum</th><th>Two-step</th></tr>
<tr>
<td></td>
<td>eywords over Shakespeare's plays in XML, the user might be interested in knowing how the keywords cooccur. In this paper, we focus on XML trees and define XML keyword, proximity queries to return the (possibly heterogeneous) set of minimum connecting trees (MCTs) of the matches to the individual keywords in the query. We consider efficiently executing keyword proximity queries on labeled trees (XML) in various settings: 1) when the XML database has been preprocessed and 2) when no indices are available on the XML database. We perform a detailed experimental evaluation to study the benefits of our approach and show that our algorithms considerably outperform prior algorithms and other applicable approaches.
(3):
We consider the problem of efficiently producing ranked results for keyword search queries over hyperlinked XML documents. Evaluating keyword search queries over hierarchical XML documents, as opposed to (conceptually) flat HTML documents, introduces many new challenges. First, XML keyword search queries do not always return entire documents, but can return deeply nested XML elements that contain the desired keywords. Second, the nested structure of XML implies that the notion of ranking is no longer at the granularity of a document, but at the granularity of an XML element. Finally, the notion of keyword proximity is more complex in the hierarchical XML data model. In this paper, we present the XRANK system that is designed to handle these novel features of XML keyword search. Our experimental results show that XRANK offers both space and performance benefits when compared with existing approaches. An interesting feature of XRANK is that it naturally generalizes a hyperlink based HTML search engine such as Google. XRANK can thus be used to query a mix of HTML and XML documents.</td>
<td></td>
<td>n to anybody. As the size of XML sources grows rapidly, the need for efficient search indices on XML data that support keyword search increases. In this paper, we present an approach of XML keyword search which is based on the DAG of the XML data, where repeated substructures are considered only once, and therefore, have to be searched only once. As our performance evaluation shows, this DAG-based extension of the set intersection search algorithm[1], [2], can lead to search times that are on large documents more than twice as fast as the search times of the XML-based approach. Additionally, we utilize a smaller index, i.e., we consume less main memory to compute the results.|||||Recent works have shown the benefits of keyword proximity search in querying XML documents in addition to text documents.
Comment:
- Copied the first tokens only
- No indication of MDS</td>
<td>we present an approach of XML keyword search which is based on the set intersection search algorithm, where repeated substructures are considered only once, and therefore, have to be searched only once.
we perform a detailed experimental evaluation to study the benefits of our approach and show that our algorithms considerably outperform prior algorithms and other applicable approaches.
Comment:
- Very weak signs of MDS, with all 3 articles covered.
- However, they are all presented as a single document.
- The model also performs like an extractive model</td>
<td>t HTML documents, introduces many new challenges," the authors write. "First, Xie XRANK does not always return entire documents, but can return deeply nesting markup that contains the desired keywords. Second, the nesting structure of markup implies that the notion of ranking is no longer at the granularity of a document, but at the granular of an markup. Finally, the notion of keyword proximity is more complex in the hierarchical markup data model. In this paper, we present the XRANK system that is designed to handle these novel features of Xie keyword search. Our experimental results show that XRANK offers both space and performance benefits when compared with existing approaches. An interesting feature of XRANK is that it naturally generalizes a hyperlink based HTML search engine such as Google. XRANK can thus be used to query a mix of HTML and markup markup."
Comment:
- Copied the 3rd article in full
- No indication of MDS</td>
<td>ch engine that can be used to query a mix of hyperlink based XML documents. X-RANK is designed to handle the problem of efficiently producing ranked results for keyword search queries over hyperlinked XML documents, as opposed to (conceptually) flat HTML documents.
Comment:
- Shows very weak signs of MDS, with an attempt to reference multiple @cite.
- However, none of the sentences generated are coherent, showing difficulties of the model when dealing with articles filled with acronyms (XML, HTML) which it does not understand</td>
<td>Comment:
- Shows very weak signs of MDS, with an attempt to reference multiple @cite.
- Even worse than tuned LED
- However, none of the sentences generated are coherent, showing difficulties of the model when dealing with articles filled with acronyms (XML, HTML) which it does not understand</td>
<td>der the nested structure of the XML documents. XRANK can be used to query a mix of HTML and XML documents, but it does not require a hyperlink search engine, and it does not need a hyperlink, and it is not required to search for hyperlinks in a hyperlink document. Moreover, XRANK is designed to handle the nested structures of the XML document. In contrast, our approach is based on a DAG-based extension of the set intersection search algorithm, which can lead to search times that are more than twice as fast as the search time of the XML-based approach.
Comment:
- Shows signs of MDS, with the yellow part contrasting the approach of the main article with that of the 3rd (blue)
- However, the red parts still show hallucination
- The first step LED model in the two-step model probably helped simplify the meaning of the individual passages?</td>
</tr>
<tr>
<td>4858</td>
<td>(1):
It is shown that one can count @math -edge paths in an @math -vertex graph and @math -set @math -packings on an @math -element universe, respectively, in time @math and @math , up to a factor polynomial in @math , @math , and @math ; in polynomial space, the bounds hold if multiplied by @math or @math , respectively. These are implications of a more general result: given two set families on an @math -element universe, one can count the disjoint pairs of sets in the Cartesian product of the two families with @math basic operations, where @math is the number of members in the two families and their subsets.
(2):
We present a fast algorithm for the subset convolution problem:given functions f and g defined on the lattice of subsets of ann-element set n, compute their subset convolution f*g, defined for S⊆ N by [(f * g)(S) = [T ⊆ S] f(T) g(S T),,] where addition and multiplication is carried out in an arbitrary ring. Via Mobius transform and inversion, our algorithm evaluates the subset convolution in O(n2 2n) additions and multiplications, substanti y improving upon the straightforward O(3n) algorithm. Specifically, if the input functions have aninteger range [-M,-M+1,...,M], their subset convolution over the ordinary sum--product ring can be computed in O(2n log M) time; the notation O suppresses polylogarithmic factors.Furthermore, using a standard embedding technique we can compute the subset convolution over the max--sum or min--sum semiring in O(2n M) time. To demonstrate the applicability of fast subset convolution, wepresent the first O(2k n2 + n m) algorithm for the Steiner tree problem in graphs with n vertices, k terminals, and m edges with bounded integer weights, improving upon the O(3kn + 2k n2 + n m) time bound of the classical Dreyfus-Wagner algorithm. We also discuss extensions to recent O(2n)-time algorithms for covering and partitioning problems (Bjorklund and Husfeldt, FOCS 2006; Koivisto, FOCS 2006).
(3):
Given a set @math with @math elements and a family @math of subsets, we show how to partition @math into @math such subsets in @math time. We also consider variations of this problem where the subsets may overlap or are weighted, and we solve the decision, counting, summation, and optimization versions of these problems. Our algorithms are based on the principle of inclusion-exclusion and the zeta transform. In effect we get exact algorithms in @math time for several well-studied partition problems including domatic number, chromatic number, maximum @math -cut, bin packing, list coloring, and the chromatic polynomial. We also have applications to Bayesian learning with decision graphs and to model-based data clustering. If only polynomial space is available, our algorithms run in time @math if membership in @math can be decided in polynomial time. We solve chromatic number in @math time and domatic number in @math time. Finally, we present a family of polynomia</td>
<td>Concerning set packings the situation is analogous, albeit the research has been somewhat less extensive. Deciding whether a given family of @math subsets of an @math -element universe contains a @math -packing is known to be W[1]-hard @cite , and thus it is unlikely that the problem is fixed parameter tractable, that is, solvable in time @math for some function @math and constant @math . If @math is fairly large, say exponential in @math , the fastest known algorithms actually count the packings by employing the inclusion--exclusion machinery @cite @cite and run in time @math . This bound holds also for the presented algorithm (cf. Theorem).</td>
<td>Rouge scores:
- Rouge 2:
0.1079 (prec)
0.1442 (recall)
0.1235 (f-1)
- Rouge L:
0.1714 (prec)
0.2286 (recall)
0.1959 (f-1)
Summary:
It is shown that one can count @math -edge paths in a @math -vertex graph and @math -set @math -packings on an @math -element universe, respectively, in time @math and @math, up to a factor polynomial in @math, @math, and @math ; in polynomial space, the bounds hold if multiplied by @math or @math, respectively. These are implications of a more general result: given two set families on an @x -element universe, one can count the disjoint pairs of sets in the Cartesian product of the two families with @math basic operations, where @math is the number of members in the two families and their subsets.|||||We present a fast algorithm for the subset convolution problem:given functions f and g defined on the lattice of subsets of ann-element set n, compute their subset convolution f*g, defined for S ⊆
Comment:
- Copied the first tokens only
- No indication of MDS
- Random word change noted (red)</td>
<td>Rouge scores:
- Rouge 2:
0.0612 (prec)
0.0577 (recall)
0.0594 (f-1)
- Rouge L:
0.2222 (prec)
0.2095 (recall)
0.2157 (f-1)
Summary:
given a set @math with @math elements and a family @math of subsets, we show how to partition @math into @math such subsets in @math time.
we also consider variations of this problem where the subsets may overlap or are weighted, and we solve the decision, counting, summation, and optimization versions of these problems.
our algorithms are based on the principle of inclusion-exclusion and the zeta transform.
in effect we get exact algorithms in time @math for several well - studied partition problems including domatic number, chromatic number, maximum @math -cut, bin packing, list coloring, and the chromatic polynomial.
Comment:
- Copied all of the 3rd article except for last 2 sentences
- No indication of MDS</td>
<td>Rouge scores:
- Rouge 2:
0.0791 (prec)
0.1346 (recall)
0.0996 (f-1)
- Rouge L:
0.1348 (prec)
0.2286 (recall)
0.1696 (f-1)
Summary:
It is shown that one can count @math -edge paths in an @math -vertex graph and @math -set @math -packings on an @math -element universe, respectively, in time @math and @math, up to a factor polyn Chocobo in @ math, @math, and @math ; in polyn Chocobo space, the bounds hold if multiplied by @math or @math, respectively. These are implications of a more general result: given two set families on an @ math -element universe, one can count the disjoint pairs of sets in the Cartesian product of the two families with @math basic operations, where @math is the number of members in the two families and their subsets. We present a fast algorithm for the subset convolution problem:given functions f and g defined on the lattice of subsets of ann-element set n, compute their subset convolution f*g, defined for S⊆ N by [(f * g)(S) = [T ⊆ S] f(T) g(S T),]where addition and multiplication is carried out in an arbitrary ring. Via Mobius transform and inversion, our algorithm evaluates the subset convolved in O(</td>
<td>Rouge scores:
- Rouge 2:
0.125 (prec)
0.0769 (recall)
0.0952 (f-1)
- Rouge L:
0.2923 (prec)
0.181 (recall)
0.2235 (f-1)
Summary:
For the Steiner tree problem @ cite, one can count the disjoint pairs of sets in the Cartesian product of the two families with @ math basic operations, where @math is the number of members in the two families and their subsets @math. For the subset convolution problem, one can count @math -edge paths in an @math -vertex graph and @math -set @math -packings in time @math.
Comment:
- No indication of MDS behavior, basically just mixing in some words of the 2nd article (green) with the 1st (yellow)
- Shows once again how the model fails to understand words used in unusual settings (e.g. mathematics)</td>
<td>Rouge scores:
- Rouge 2:
0.082 (prec)
0.0962 (recall)
0.0885 (f-1)
- Rouge L:
0.1951 (prec)
0.2286 (recall)
0.2105 (f-1)
Summary:
In @cite, Bjorklund and Husfeldt showed that one can partition a set @math into @math subsets in @math time, where @math is the number of members in the two families and their subsets. They also showed how to partition the set into @math such subsets in time @math, where @math and @math are the subsets of @math. They also gave a family of polynomial space approximation algorithms that find a number between @math and the @math in @math. @cite gave a fast algorithm for the subset convolution problem in graphs with @math vertices, @math terminals, and @math edges with bounded integer weights, improving upon the classical Dreyfus-Wagner algorithm. They also discussed extensions to recent @math -time algorithms for covering and partitioning problems.
Comment:
- Some weak signs of MDS, with the 2nd (green) and 3rd (blue) articles referenced and mixed with the 1st (yellow)
- Shows once again how the model fails to understand</td>
<td>Rouge scores:
- Rouge 2:
0.0745 (prec)
0.0673 (recall)
0.0707 (f-1)
- Rouge L:
0.1895 (prec)
0.1714 (recall)
0.18 (f-1)
Summary:
In @cite, Dreyfus-Wagner algorithm @cite was presented, which is based on an @math time bound of the classical DREW algorithm. The algorithm was extended to the Steiner Tree problem by @cite and presented an O(3n - @math) time bound for the subgraphs with bounded integer weights. In the case where @math is not available, @cite showed how to partition @math into @math such subsets in polynomial time and @math in time @cite. In this case, @math and @math are polynomial in time, and @cite gives an approximation algorithm for chromatic number and domatic number.
Comment:
- Best performing of all 3 tuned models, showing strong signs of MDS with the 2nd (green) and 3rd (blue) articles reflected.
- The red part appears to contain some factual errors, with the original "O(3kn + 2k n2 + n m)" changed to "O(3n - @math)"
- The first step LED model in the two-step model probably</td>
</tr>
</table>

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
	l space approximation algorithms that find a number between @math and @ math in time @math .				**Comment:** - Copied the first tokens only - No indication of MDS - Random word change noted (red)		words used in unusual settings (e.g. mathematics)	helped simplify the meaning of the individual passages?
5068	**(1):** Convolutional neural networks have gained a remarkable success in computer vision. However, most usable network architectures are hand-crafted and us ually require expertise and elaborate design. In this paper, we provide a block -wise network generation pipeline called BlockQNN which automatically build s high-performance networks using the Q-Learning paradigm with epsilon-gre edy exploration strategy. The optimal network block is constructed by the lea rning agent which is trained sequentially to choose component layers. We sta ck the block to construct the whole auto-generated network. To accelerate th e generation process, we also propose a distributed asynchronous framework and an early stop strategy. The block-wise generation brings unique advantag es: (1) it performs competitive results in comparison to the hand-crafted state -of-the-art networks on image classification, additionally, the best network ge nerated by BlockQNN achieves 3.54 top-1 error rate on CIFAR-10 which beats all existing auto-generate networks. (2) in the meanwhile, it offers tremendou s reduction of the search space in designing networks which only spends 3 da ys with 32 GPUs, and (3) moreover, it has strong generalizability that the net work built on CIFAR also performs well on a larger-scale ImageNet dataset. **(2):** This paper introduces the application of gradient descent methods to meta-le arning. The concept of "meta-learning", i.e. of a system that improves or disc overs a learning algorithm, has been of interest in machine learning for decad es because of its appealing applications. Previous meta-learning approaches h ave been based on evolutionary methods and, therefore, have been restricte d to small models with few free parameters. We make meta-learning in large systems feasible by using recurrent neural networks withth eir attendant lear ning routines as meta-learning systems. Our system derived complex well per forming learning algorithms from scratch. In this paper we also show that our approachp erforms non-stationary time series prediction. **(3):** Different researchers hold different views of what the term meta-learning exa ctly means. The first part of this paper provides our own perspective view in which the goal is to build self-adaptive learners (i.e. learning algorithms that i mprove their bias dynamically through experience by accumulating meta-kno wledge). The second part provides a survey of meta-learning as reported by t he machine-learning literature. We find that, despite different views and rese arch lines, a question remains constant: how can we exploit knowledge about learning (i.e. meta-knowledge) to improve the performance of learning algori thms? Clearly the answer to this question is key to the advancement of the fie ld and continues being the subject of intensive research. **(4):** The move from hand-designed features to learned features in machine learni ng has been wildly successful. In spite of this, optimization algorithms are still designed by hand. In this paper we show how the design of an optimization al gorithm can be cast as a learning problem, allowing the algorithm to learn to exploit structure in the problems of interest in an automatic way. Our learned algorithms, implemented by LSTMs, outperform generic, hand-designed com petitors on the tasks for which they are trained, and also generalize well to ne w tasks with similar structure. We demonstrate this on a number of tasks, incl uding simple convex problems, training neural networks, and styling images w ith neural art.	Another bunch of related works i nclude hyper-parameter optimiz ation @cite , meta-learning @cit e and learning to learn methods @cite @cite . However, the goal of these works is to use meta-da ta to improve the performance o f the existing algorithms, such as finding the optimal learning rate of optimization methods or the o ptimal number of hidden layers t o construct the network. In this paper, we focus on learning the entire topological architecture of network blocks to improve the p erformance.	**Rouge scores:** - Rouge 2: 0.0596 (prec) 0.1169 (recall) 0.0789 (f-1) - Rouge L: 0.1053 (prec) 0.2051 (recall) 0.1391 (f-1) **Summary:** Convolutional neural networks have gained a remarkable succe ss in computer vision. However, most usable network architect ures are hand-crafted and usual ly require expertise and elabora te design. In this paper, we prov ide a block-wise network gener ation pipeline called BlockQNN which automatically builds high -performance networks using th e Q-Learning paradigm with eps ilon-greedy exploration strateg y. The optimal network block is constructed by the learning age nt which is trained sequentially to choose component layers. W e stack the block to construct th e whole auto-generated networ k. To accelerate the generation process, we also propose a distr ibuted asynchronous framewor k and an early stop strategy. Th e block-wise generation brings unique advantages: (1) it perfor ms competitive results in comp arison to the hand-crafted state -of-the-art networks on image c lassification, additionally, the b est network generated by Block QNN achieves 3.54 top-1 error r ate on CIFAR-10 which beats all existing auto-generate network s. (2) **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0723 (prec) 0.0779 (recall) 0.075 (f-1) - Rouge L: 0.1548 (prec) 0.1667 (recall) 0.1605 (f-1) **Summary:** in this paper we show how the design of an optimization algori thm can be cast as a learning pr oblem, allowing the algorithm t o learn to exploit structure in th e problems of interest in an aut omatic way. we make meta-learning in large systems feasible by using recur rent neural networks withth eir attendant learning routines as meta-learning systems. our learned algorithms outperf orm generic, hand-designed co mpetitors on the tasks for whic h they are trained, and also gen eralize well to new tasks with si milar structure. **Comment:** - Copied most of the 4th article (purple) and just inserted a sentence from the 2nd (green) - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0513 (prec) 0.1299 (recall) 0.0735 (f-1) - Rouge L: 0.0969 (prec) 0.2436 (recall) 0.1387 (f-1) **Summary:** Convolutional neural networks have gained a remarkable succe ss in computer vision. However, most usable network architect ures are hand-crafted and usual ly require expertise and elabora te design. In this paper, we prov ide a block-wise network gener ation pipeline called BlockQNN which automatically builds high-performance networks using th e Q-learning paradigm with epsi lon-greedy exploration strategy. The optimal network block is c onstructed by the learning agen t which is trained sequentially t o choose component layers. We stack the block to construct the whole auto-generated networ k. To accelerate the generation process, we also propose a distr ibuted asynchronous framewor k and an early stop strategy. Th e block-wise generation brings unique advantages: (1) it perfor ms competitive results in comp arison to the hand-crafted state -of-the-art networks on image c lassification, additionally, the be st network generated by BlockQ NN achieves 3.54 top-1 error ra te on CIFAR-10 which beats all e xisting auto-generate networks. (2) in the meanwhile, it offers t remendous reduction of the sea rch space in designing networks which only spends 3 days with 32 GPUs, and (3) moreover, it h as strong generalizability that th e network built on CIFar also pe rforms well on a larger-scale Im ageNet dataset. **Comment:** - Copied the first tokens only - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0811 (prec) 0.039 (recall) 0.0526 (f-1) - Rouge L: 0.1842 (prec) 0.0897 (recall) 0.1207 (f-1) **Summary:** In this paper, we propose a bloc k-wise network generation pipel ine using the Q-learning paradig m with epsilon-greedy explorati on strategy. We also propose a distributed asynchronous frame work and an early stop strategy to accelerate the generation pr ocess. **Comment:** - Only extracted 2 sentences from the 1st article (yellow) - No indication of MDS	**Rouge scores:** - Rouge 2: 0.0952 (prec) 0.1299 (recall) 0.1099 (f-1) - Rouge L: 0.1698 (prec) 0.2308 (recall) 0.1957 (f-1) **Summary:** Recently, meta-learning @cite @cite has gained a lot of attenti on in the machine learning liter ature. In this paper, we propose a block-wise network generatio n pipeline called BlockQNN whic h automatically builds high-perf ormance networks using the Q-Learning paradigm with epsilon-greedy exploration strategy. Th e optimal network block is cons tructed sequentially by the lear ning agent which is trained sequ entially to choose component la yers. In addition, it offers treme ndous reduction of the search s pace in designing networks whi ch only spend 3 days with 32 GP Us, and it has strong generaliza bility that the network built on CIFAR-10 also performs well on a larger-scale ImageNet datase t. **Comment:** - No indication of MDS, except for the red part where the model noted "meta-learning" is a recurring theme	**Rouge scores:** - Rouge 2: 0.0899 (prec) 0.1039 (recall) 0.0964 (f-1) - Rouge L: 0.2111 (prec) 0.2436 (recall) 0.2262 (f-1) **Summary:** @cite @cite proposed a gradien t descent method for meta-lear ning. The gradient descent met hod is similar to our method, bu t differs from our method in tw o important aspects. First, the g radient descent method does n ot rely on the gradient of the gr adient. Second, it does not requ ire the gradient of a gradient to be learned. In this paper, we sh ow that the gradient of our met hod can be used as a meta-lear ning method. We also show tha t our approach can be applied t o large systems with few free p arameters. **Comment:** - Shows some sign of MDS, with the 2nd article (green) being mentioned. - However, the other parts are mostly hallucinations

(c) Long samples

- 485: sample where the strong performance of the tuned and two-step models can be seen

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
485	**(1):** Learning from web data has attracted lots of research interest in recent years. However, crawled web images usually have two types of noises, label noise a nd background noise, which induce extra difficulties in utilizing them effective ly. Most existing methods either rely on human supervision or ignore the back ground noise. In this paper, we propose a novel method, which is capable of h	In learning classifier with web da ta, previous works focus on hand ling the label noise in three direc tions, removing label noise @cit e @cite @cite @cite @cite @cit e @cite , building noise-robust m	**Rouge scores:** - Rouge 2: 0.0175 (prec) 0.0732 (recall) 0.0283 (f-1) - Rouge L:	**Rouge scores:** - Rouge 2: 0.0258 (prec) 0.0976 (recall) 0.0408 (f-1) - Rouge L:	**Rouge scores:** - Rouge 2: 0.0048 (prec) 0.0244 (recall) 0.0081 (f-1) - Rouge L:	**Rouge scores:** - Rouge 2: 0.0485 (prec) 0.122 (recall) 0.0694 (f-1) - Rouge L:	**Rouge scores:** - Rouge 2: 0.0266 (prec) 0.122 (recall) 0.0437 (f-1) - Rouge L:	**Rouge scores:** - Rouge 2: 0.0287 (prec) 0.1463 (recall) 0.048 (f-1) - Rouge L:

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

No.	Abstracts	Label	Base LED	Off-the-shelf LED	Off-the-shelf Centrum	Tuned LED	Tuned Centrum	Two-step
	nd traditional supervised learning is still very large, due to the label noise of w eb data as well as the domain shift between web data and test data. To fill thi s gap, most existing methods propose to purify or augment web data using in stance-level supervision, which generally requires heavy annotation. Instead, we propose to address the label noise and domain shift by using more accessi ble category-level supervision. In particular, we build our deep probabilistic fr amework upon variational autoencoder (VAE), in which classification network and VAE can jointly leverage category-level hybrid information. Then, we ext end our method for domain adaptation followed by our low-rank refinement strategy. Extensive experiments on three benchmark datasets demonstrate th e effectiveness of our proposed method. **(7):** Label noise is an important issue in classification, with many potential negativ e consequences. For example, the accuracy of predictions may decrease, whe reas the complexity of inferred models and the number of necessary training samples may increase. Many works in the literature have been devoted to the study of label noise and the development of techniques to deal with label noi se. However, the field lacks a comprehensive survey on the different types of label noise, their consequences and the algorithms that consider label noise. This paper proposes to fill this gap. First, the definitions and sources of label n oise are considered and a taxonomy of the types of label noise is proposed. S econd, the potential consequences of label noise are discussed. Third, label n oise-robust, label noise cleansing, and label noise-tolerant algorithms are revi ewed. For each category of approaches, a short discussion is proposed to help the practitioner to choose the most suitable technique in its own particular fi eld of application. Eventually, the design of experiments is also discussed, wh at may interest the researchers who would like to test their own algorithms. I n this paper, label noise consists of mislabeled instances: no additional inform ation is assumed to be available like e.g., confidences on labels. **(8):** In this paper, we study a classification problem in which sample labels are ran domly corrupted. In this scenario, there is an unobservable sample with noise -free labels. However, before being observed, the true labels are independent ly flipped with a probability @math , and the random label noise can be class-conditional. Here, we address two fundamental problems raised by this scena rio. The first is how to best use the abundant surrogate loss functions designe d for the traditional classification problem when there is label noise. We prov e that any surrogate loss function can be used for classification with noisy lab els by using importance reweighting, with consistency assurance that the labe l noise does not ultimately hinder the search for the optimal classifier of the n oise-free sample. The other is the open problem of how to obtain the noise ra te @math . We show that the rate is upper bounded by the conditional proba bility @math of the noisy sample. Consequently, the rate can be estimated, b ecause the upper bound can be easily reached in classification problems. Exp erimental results on synthetic and real datasets confirm the efficiency of our methods. **(9):** Current approaches for fine-grained recognition do the following: First, recrui t experts to annotate a dataset of images, optionally also collecting more stru ctured data in the form of part annotations and bounding boxes. Second, trai n a model utilizing this data. Toward the goal of solving fine-grained recogniti on, we introduce an alternative approach, leveraging free, noisy data from th e web and simple, generic methods of recognition. This approach has benefits in both performance and scalability. We demonstrate its efficacy on four fine -grained datasets, greatly exceeding existing state of the art without the man ual collection of even a single label, and furthermore show first results at scali ng to more than 10,000 fine-grained categories. Quantitatively, we achieve to p-1 accuracies of (92.3 ,) on CUB-200-2011, (85.4 ,) on Birdsnap, (93.4 ,) on F GVC-Aircraft, and (80.8 ,) on Stanford Dogs without using their annotated trai ning sets. We compare our approach to an active learning approach for expan ding fine-grained datasets.							