## A CONTRIBUTION DISTRIBUTION IN THE PAPERS

Different writing styles might locate and express contributions in different ways. To understand the global tendency of contribution locations in a paper, we take each sentence from the paper texts themselves in the training set and annotated contributions using the learned sentence classifier. We then group them into 10 bins according to the relative location of the sentences in the papers they belong to and constructed a distribution which summarizes the proportion of sentences labeled as contributions in each bin. Fig 4 shows the percentages of such sentences for each bin. The graph shows that no bin positions in the papers tend to describe contributions more than 50% of the time. Surprisingly, the first 10% of the papers have the lowest chance of describing the contributions, which is counter-intuitive to the general idea that papers tend to discuss the introduction and highlights of the paper at the beginning.

Figure 4: Frequency of contribution mentions in different parts of papers. Around 22% of sentences in the first bin are labeled as "contribution".
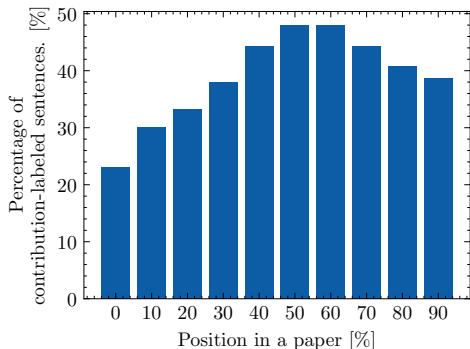
B DOMAIN DISTRIBUTION

We show the Top-10 domain distribution of our dataset in Figure 5. Biology and Medicine are the two most dominant domains. 5th most frequent "domain" is indicated as N/A, meaning that the domain information was not available by S2ORC. Qualitatively, papers in the Biology domain tend to have a similarly formatted summary style to that of Medicine.

Figure 5: Top-10 frequent domains in our dataset.
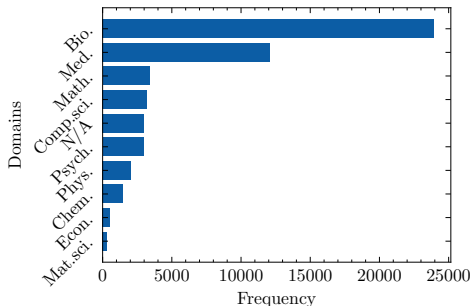
C HUMAN EVALUATION OF DISENTANGLEMENT

In addition to various automatic evaluation, we perform human evaluation on disentanglement to understand which models human annotators prefer. We use Best-Worst scaling (Kiritchenko & Mohammad, 2017) over the 4-tuples of summaries on the 50 random samples from the test set and have 3 annotators pick the best and the worst contribution and context summary pairs in terms of disentanglement. The rating in Table 7 shows the percentage a model is chosen as the best minus the percentage a model is chosen as the worst; rating ranges from -1 to 1. A similar trend to automatic disentanglement evaluation is observed here as well in that introducing the informativeness objective significantly improves MULTIHEAD model, while it conversely affects CONTROLCODE model.

Table 7: Disentanglement using Best-Worst scaling.

Model	Rating
CC	**0.027**
CC+INF	0.020
MH	-0.073
MH+INF	**0.027**

D GENERATED FULL SAMPLES FROM DIFFERENT MODELS

We show additional samples generated from CONTROLCODE model in Table 8, 9, and 10. A common failure case of all our models is the lack of disentanglement between the two summaries. While samples start generating on-topic, the model for context summary generation starts generating the contributions at one point. This phenomenon can conspicuously be observed in Table 8 and 10.

Table 8: Generated Sample. In this example, the red highlighted segments describe the contribution.

Original Abstract: Domain adaption (DA) allows machine learning methods trained on data sampled from one distribution to be applied to data sampled from another. It is thus of great practical importance to the application of such methods. Despite the fact that tensor representations are widely used in Computer Vision to capture multi-linear relationships that affect the data, most existing DA methods are applicable to vectors only. This renders them incapable of reflecting and preserving important structure in many problems. We thus propose here a learning-based method to adapt the source and target tensor representations directly, without vectorization. In particular, a set of alignment matrices is introduced to align the tensor representations from both domains into the invariant tensor subspace. These alignment matrices and the tensor subspace are modeled as a joint optimization problem and can be learned adaptively from the data using the proposed alternative minimization scheme. Extensive experiments show that our approach is capable of preserving the discriminative power of the source domain, of resisting the effects of label noise, and works effectively for small sample sizes, and even one-shot DA. We show that our method outperforms the state-of-the-art on the task of cross-domain visual recognition in both efficacy and efficiency, and particularly that it outperforms all comparators when applied to DA of the convolutional activations of deep convolutional networks.

Generated Abstract: The difficulty of securing an appropriate and exhaustive set of training data, and the tendency for the domain of application to drift over time, often lead to variations between the distributions of the training (source) and test (target) data. In Machine Learning this problem is labeled domain mismatch. Failing to model such a distribution shift may cause significant performance degradation. Domain adaptation (DA) techniques capable of addressing this problem of distribution shift have thus received significant attention recently [24]. The assumption underpinning DA is that, although the domains differ, there is sufficient commonality to support adaptation. Many approaches have modeled this commonality by learning an invariant subspace, or set of subspaces. These methods are applicable to vector data only, however. Applying these methods to structured high-dimensional representations (e.g., convolutional activations), thus requires that the data be vectorized first. Although this solves the algebraic issue, it does not solve the underlying problem. Tensor

Contribution: To address these issues, we propose a novel approach termed Tensor-Aligned Invariant Subspace Learning (TAISL) to learn an invariant tensor subspace that is able to adapt the tensor representations directly. By introducing a set of alignment matrices, the tensors from the source domain are aligned to an underlying tensor space shared by the target domain. Instead of executing a holistic adaptation (where all feature dimensions would be taken into account), our approach performs mode-wise partial adaptation where each mode is adapted separately to avoid the curse of dimensionality. We also propose an alternating minimization scheme which allows the problem to be effectively optimized by off-the-shelf solvers. Extensive experiments on cross-domain visual recognition demonstrate the following merits of our approach: i) it effectively reduces the domain discrepancy and preserves the discriminative power of the original representations; ii) it is applicable to small sample size adaptation, even when there is only one source

Context: Deep convolutional neural networks (CNNs) represent the state-of-the-art method for a substantial number of visual tasks. The activations of such CNNs, and the interactions between them, are naturally represented as tensors, meaning that DA should also be applied using this representation. However, after vectorization, many existing approaches become sensitive to the scarcity of source data (compared to the number of dimensions) and noise in the labels. The proposed direct tensor method uses much lower dimensional entities, thus avoiding these estimation problems. To address these issues we propose to learn an invariant tensor subspace that is able to adapt the tensor representations directly. We show in Section 5 that the proposed method outperforms all comparators in DA of the Convolutional Activation of CNNs. Higher-order tensor modeling offers us an opportunity to investigate multiple interactions and couplings that capture the commonality and differences between domains. Following this idea, a novel approach

Table 9: Generated Sample. Context summary has a slight erroneous decoding of contribution information.

Original Abstract: Conventional wisdom holds that model-based planning is a powerful approach to sequential decision-making. It is often very challenging in practice, however, because while a model can be used to evaluate a plan, it does not prescribe how to construct a plan. Here we introduce theÏmagination-based Planner, the first model-based, sequential decision-making agent that can learn to construct, evaluate, and execute plans. Before any action, it can perform a variable number of imagination steps, which involve proposing an imagined action and evaluating it with its model-based imagination. All imagined actions and outcomes are aggregated, iteratively, into aplan contextẅhich conditions future real and imagined actions. The agent can even decide how to imagine: testing out alternative imagined actions, chaining sequences of actions together, or building a more compleximagination treeby navigating flexibly among the previously imagined states using a learned policy. And our agent can learn to plan economically, jointly optimizing for external rewards and computational costs associated with using its imagination. We show that our architecture can learn to solve a challenging continuous control problem, and also learn elaborate planning strategies in a discrete maze-solving task. Our work opens a new direction toward learning the components of a model-based planning system and how to use them.

Generated Abstract: The Imagination-based Planner (IBP) is a model-based agent which learns from experience all aspects of the planning process: how to construct, evaluate, and execute a plan. The IBP learns when to act versus when to imagine, and if imagining, how to select states and actions for aggregating the proposed actions and evaluations into a useful plan. Through training, it effectively develops a planning algorithm tailored to the target problem. The learned algorithm allows it to flexibly explore, and exploit regularities in, the state and action spaces. In two experiments we evaluated a continuous IBP implementation on a challenging continuous control task, and a discrete IBP in a maze-solving problem. A fully learnable plan-based planning agent for continuous control. An agent that learns to construct a plan via model-inspired imagination was trained to perform čonditionaländ ädaptive computation,ẅhich results in a dynamic computational graph recently [7] trained neural networks trained to

Contribution: In this paper we introduce the Imagination-based Planner (IBP), a model-based planning agent which learns from experience all aspects of the planning process: how to construct, evaluate, and execute a plan. The IBP learns when to act versus when to imagine, and if imagining, how to select states and actions to evaluate which will help minimize its external task loss and internal resource costs. Through training, it effectively develops a planning algorithm tailored to the target problem. In two experiments we evaluated a continuous IBP implementation on a challenging continuous control task, and a discrete IBP in a maze-solving problem.

Context: Model-based planning involves proposing sequences of actions, evaluating them under a model of the world, and refining these proposals to optimize expected rewards. Several key advantages of model-free methods are that models support generalization to states not previously experienced, help express the relationship between present actions and future rewards, and can resolve states which are aliased in value-based approximations. These advantages are especially pronounced in problems with complex and stochastic environmental dynamics, sparse reward, and restricted trial-and-error experience. Yet even with an accurate model, planning is often very challenging because while a model can be used to evaluate a plan, it does not prescribe how to construct a plan. Existing techniques for model based planning are most effective in small-scale problems, often require background knowledge of the domain, and use pre-defined solution strategies. Here we introduce the Imagination-based Planner (IBP), a model-based agent which learns from experience all aspects of the planning

Table 10: Generated Sample. More than half of the context summary is describing contribution.

Original Abstract: We provide a generalization of discriminative lexicalized shift reduce parsing techniques for phrase structure grammar to a wide range of morphologically rich languages. The model is efficient and outperforms recent strong baselines on almost all languages considered. It takes advantage of a dependency based modelling of morphology and a shallow modelling of constituency boundaries.

Generated Abstract: This paper shows that an efficient lexicalized phrase structure parser -modelling both dependencies and morphology -already significantly improves parsing accuracy. But we also show that an additional modelling of spans and constituency provides additional robustness that contributes to yield state of the art results on almost all languages considered, while remaining quite quite efficient. Moreover given the availability of existing multi-view treebanks (Bhatt et al., 2009; Seddah et al. 2014), our proposed solution only requires a lightweight infrastructure to achieve multilin-gual parsing without requiring costly languaged dependent modifications such as feature engineering. This method was used for the construction of the dataset for the SPMRL 2013 shared task. We first review the properties of multiview treebanks for a number of morphologically rich languages for which they are annotated both for constituents and dependencies that have the property to be token-wise aligned. These double annotations are typically obtained by converting a constituency or dependency annotation into the other annotation

Contribution: In this paper, we propose a lexicalized phrase structure parsing model that can effectively take morphology into account. We show that an efficient lexicalization of both dependencies and morphology significantly improves multilingual state-of-the art results on almost all languages considered, while remaining quite efficient. Moreover, given the availability of existing multi-view treebanks, our proposed solution only requires a lightweight infrastructure to achieve multilin-gual parsing without requiring costly feature engineering such as feature engineering.

Context: Most state of the art multilingual parsers are weighted by discriminative models. Most state-of-the-art multilingual parsing methods rely on lexicalized phrase structure parsing techniques, which have recently been shown to improve performance in a variety of languages including free word order languages like English or Chinese. In this paper we show that an efficient lexicallyized parser - modelling both dependencies and morphology - already significantly improves parsing accuracy. But it also shows that an additional modelling of spans and constituency provides additional robustness that contributes to yield state-ofthe-art results on almost all languages considered, while remaining quite efficient. Moreover, given the availability of existing multi-view treebanks (Bhatt et al., 2009; Seddah et al. 2013), our proposed solution only requires a lightweight infrastructure to achieve multilin-gual parsing without requiring costly feature engineering such as feature engineering.