Domain generalization using gaussian process latent variable model

Anonymous Author(s)
Affiliation
Address
email

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

There is a new style file for papers submitted in 2016!

○ Domain generalization とは

○先行研究

`https://cmt.research.microsoft.com/NIPS2016/`

Please read carefully the instructions below and follow them faithfully.

1.1 Style

Papers to be submitted to NIPS 2016 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Since 2009 an additional ninth page *containing only acknowledgments and/or cited references* is allowed. Papers that exceed nine pages will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2016 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper compared to earlier years.

Authors are required to use the NIPS LaTeX style files obtainable at the NIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

1.2 Retrieval of style files

The style files for NIPS and other conference information are available on the World Wide Web at

`http://www.nips.cc/`

The file `nips_2016.pdf` contains these instructions and illustrates the various formatting requirements your NIPS paper must satisfy.

The only supported style file for NIPS 2016 is `nips_2016.sty`, rewritten for LaTeX 2_ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

Submitted to 30th Conference on Neural Information Processing Systems (NIPS 2016). Do not distribute.

The new LaTeX style file contains two optional arguments: `final`, which creates a camera-ready copy, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

At submission time, please omit the `final` option. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `nips_2016.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections **??**, **??**, and **??** below.

2 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section **??** regarding figures, tables, acknowledgments, and references.

3 Notations

$\mathcal{X}$: input space
$\mathcal{Y}$: output space
$\mathbb{P}_{XY}$: joint distribution on $\mathcal{X} \times \mathcal{Y}$
We define a $\mathbf{domain}$ to be $\mathbb{P}_{XY}$
We denote source domain as $\mathbb{P}^s_{XY}$, and test domain $\mathbb{P}^t_{XY}$,
or just simply $\mathbb{P}^s$ and $\mathbb{P}^t$.
$\mathfrak{P}_{\mathcal{X}\times\mathcal{Y}}$: the set of all domains
$\mathfrak{P}_{\mathcal{X}}$: the set of $\mathbb{P}_X$ on X
$\mathfrak{P}_{\mathcal{Y}|\mathcal{X}}$: the set of $\mathbb{P}_{Y|X}$ on Y given X
Some labeled Data $\mathcal{D}_{\mathcal{S}} = \{(x^s_i, y^s_i)^n_{i=1}\}$ are available.
Given unlabeled Data $\mathcal{D}_{\mathcal{T}} = \{(x^t_i)^m_{i=1}\}$, we want to predict the labels of them using the $\mathcal{D}_{\mathcal{S}}$.

Domains are sampled from $\mathcal{P}$ on $\mathfrak{P}_{\mathcal{X}\times\mathcal{Y}}$.
Domains are not observed directly.
Instead we observe N samples $\mathcal{S} = \{S^i\}^N_{i=1}$,
where $S^i = \{(x^i_k, y^i_k)^{n_i}_{k=1}\}$ is sampled from $\mathbb{P}^i_{XY}$ and each $\mathbb{P}^i_{XY}$ is sampled from $\mathcal{P}$.
So, $\mathcal{S}$ are not i.i.d.
$\mathcal{H}$: RKHS on $\mathcal{X}$ with kernel k : $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$
$\mathcal{F}$: RKHS on $\mathcal{Y}$ with kernel l : $\mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$
We also express:
kernel $k(\cdot, \cdot)$ induce the mapping $x \to \phi(x) \in \mathcal{H}$
kernel $l(\cdot, \cdot)$ induce the mapping $y \to \varphi(y) \in \mathcal{F}$
$\Sigma_{xx}, \Sigma_{yx}, \Sigma_{xy}, \Sigma_{yy}$: covariance operators in and between $\mathcal{H}, \mathcal{F}$

3.1 inducing points method

In the way of inducing points approximation, we will find inducing points that explain the model well by using the variational approximation. Usually, the inducing points is far smaller than the data size. Here we denote the inducing points by $\boldsymbol{Z}$, and generated t processes as U.Then, the log liklihood is written in the following way.

$$\ln P(Y|X) = \ln \int dF P(Y|F)P(F|X) = \ln \int dF dU P(Y|F)P(F,U|X,Z)P(U|Z) \tag{1}$$

$$= \ln \int dF dU Q(F,U) \frac{P(Y|F)P(F,U|X,Z)P(U|Z)}{Q(F,U)} \tag{2}$$

$$\geq \int dF dU P(F|U)Q(U) \ln \frac{P(Y|F)P(U|Z)}{Q(U)} \tag{3}$$

$$= \left\langle \ln P(Y|F) \right\rangle_{P(F|U)Q(U)} - KL\big(Q(U)|P(U|Z)\big) \tag{4}$$

Here, we denote the posterior distribution as Q(F,U). In the usual gaussian process, we parametrize the posterior Q(U) as $N(\boldsymbol{\mu}, \Sigma)$. In the inference step, by maximizing the above lower bound, we derive the variational and hyper parameters $\{\boldsymbol{\mu}, \Sigma, Z, \theta\}$ simultaneously.

The good point of this formulation is that we can reformulate this expression by the sum of data-points, therefore we can utilize the stochastic variational inference(SVI) or parallel computation.So we can apply this method to the considerably large datasets.

This kind of method can also be applied to the t process cases. Unfortunately, the analytical calculation of KL divergence is not tractable. To be more precise, we can calculate the entropy term analytically, but cannot calculate cross entropy term. The analytic form of entropy is given as follows.

We can evaluate the cross entropy by MC sampling or マクローリン expansion when the degree of freedom is large enough.

3.1.1 Latent variable model

We formulate the t process latent variable model as same as the gaussian process latent variable model(GPLVM) as follows.

We can make this latent variable model scalable .

$$\ln P(Y) \geq \int dF dU dd X P(F|U)Q(U)Q(X) \ln \frac{P(Y|F)P(U|Z)P(X)}{Q(U)Q(X)} \tag{5}$$

$$= \left\langle \ln P(Y|F) \right\rangle_{P(F|U)Q(U)Q(X)} - KL\big(Q(U)|P(U|Z)\big) - KL\big(Q(X)|P(X)\big) \tag{6}$$

Here X and P(X), Q(X) is the latent variable and its prior and posterior probability distribution. In the GPLVM, we assume that the prior distribution follows gaussian distribution $N(0, \boldsymbol{I})$, and posterior distribution $N(\boldsymbol{\mu}, diag\Sigma)$.

We can apply this scheme into the student t process. We assume that the latent probability follows どういう事前分布と事後分布をおいたのかを書く Here we assume that the prior distribution follows

3.2 Random feature sampling

3.3 Random feature sampling

4 Our Modell

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

4.1 Variational inference

4.1.1 Variational inference for supervised Model

4.2 MMD penalty for domain invariant features

We would like the latent variables to have domain invariant features. To achieve this, we add the MMD penalty to the likelihood. Using the MMD distance, we can estimate the discrepancy of distributions through the samples from those distributions very efficiently. To calculate the MMD discrepancy, we used the gaussian ARD kernel whose hyperparameters are the same for the supervised layer, and if the model is unsupervised we also learned the MMD hyperparameters with other hyperparameters. We consider two types of MMD penalty. First one measures the discrepancy between each distribution and the all combined distribution. The expression is as follows.

The second type of MMD penalty is as follows. The idea is comes from that the latent variables who have the same class have common features between domains.

After all, therefore the likelihood we calculate is as follows. The coefficient λ before MMD penalty is the regularizing parameter. We found that the accuracy of our model does not depend on this parameter so much compared to other hyperparameters or variational parameters, so we fix it to 1.0 during the all the experiment.

4.3 Adding gaussian process layers

4.4 Inference

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

5 Experiments

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

5.1 Regression

5.2 Classification

5.3 MMD penalty for domain invariant features

6 Feed forward deep student t process

$\mathcal{X}$: input space
$\mathcal{Y}$: output space
$\mathbb{P}_{XY}$: joint distribution on $\mathcal{X} \times \mathcal{Y}$
We define a **domain** to be $\mathbb{P}_{XY}$
We denote source domain as $\mathbb{P}^s_{XY}$, and test domain $\mathbb{P}^t_{XY}$,
or just simply $\mathbb{P}^s$ and $\mathbb{P}^t$.
$\mathfrak{P}_{\mathcal{X}\times\mathcal{Y}}$: the set of all domains
$\mathfrak{P}_{\mathcal{X}}$: the set of $\mathbb{P}_X$ on X
$\mathfrak{P}_{\mathcal{Y}|\mathcal{X}}$: the set of $\mathbb{P}_{Y|X}$ on Y given X
Some labeled Data $\mathcal{D}_{\mathcal{S}} = \{(x^s_i, y^s_i)^n_{i=1}\}$ are available.
Given unlabeled Data $\mathcal{D}_{\mathcal{T}} = \{(x^t_i)^m_{i=1}\}$, we want to predict the labels of them using the $\mathcal{D}_{\mathcal{S}}$.

Domains are sampled from $\mathcal{P}$ on $\mathfrak{P}_{\mathcal{X}\times\mathcal{Y}}$.
Domains are not observed directly.
Instead we observe N samples $\mathcal{S} = \{S^i\}^N_{i=1}$,
where $S^i = \{(x^i_k, y^i_k)^{n_i}_{k=1}\}$ is sampled from $\mathbb{P}^i_{XY}$ and each $\mathbb{P}^i_{XY}$ is sampled from $\mathcal{P}$.
So, $\mathcal{S}$ are not i.i.d.
$\mathcal{H}$: RKHS on $\mathcal{X}$ with kernel k : $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$\mathcal{F}$: RKHS on $\mathcal{Y}$ with kernel l : $\mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$
We also express:
kernel $k(\cdot, \cdot)$ induce the mapping $x \to \phi(x) \in \mathcal{H}$
kernel $l(\cdot, \cdot)$ induce the mapping $y \to \varphi(y) \in \mathcal{F}$
$\Sigma_{xx}, \Sigma_{yx}, \Sigma_{xy}, \Sigma_{yy}$: covariance operators in and between $\mathcal{H}, \mathcal{F}$

7 Model

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

7.1 Approximation using Random fourie feature

7.2 Approximation using inducing points

7.3 MMD penalty for domain invariant features

We would like the latent variables to have domain invariant features. To achieve this, we add the MMD penalty to the likelihood. Using the MMD distance, we can estimate the discrepancy of distributions through the samples from those distributions very efficiently. To calculate the MMD discrepancy, we used the gaussian ARD kernel whose hyperparameters are the same for the supervised layer, and if the model is unsupervised we also learned the MMD hyperparameters with other hyperparameters. We consider two types of MMD penalty. First one measures the discrepancy between each distribution and the all combined distribution. The expression is as follows.

The second type of MMD penalty is as follows. The idea is comes from that the latent variables who have the same class have common features between domains.

After all, therefore the likelihood we calculate is as follows. The coefficient λ before MMD penalty is the regularizing parameter. We found that the accuracy of our model does not depend on this parameter so much compared to other hyperparameters or variational parameters, so we fix it to 1.0 during the all the experiment.

7.4 Adding student t process layers

7.5 Inference

8 Our Model

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

8.1 Headings: second level

Second-level headings should be in 10-point type.

8.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

9 Citations, figures, tables, references

These instructions apply to everyone.

9.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `nips_2016` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{nips_2016}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

9.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

9.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

9.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table **??**.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

[1] Sample of the first footnote.

[2] As in this example.

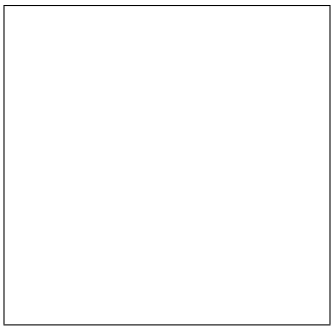

Figure 1: Sample figure caption.

Table 1: Comparison of our model and previous methods

Part		
		Size (μm)
our method	Input terminal	$\sim$100
DICA	Output terminal	$\sim$10

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table **??**.

10 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

11 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{I\!\!R} %real numbers
\newcommand{\Nat}{I\!\!N} %natural numbers
\newcommand{\CC}{I\!\!\!\!C} %complex numbers
```

Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

11.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/`

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use a ninth page as long as it contains *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.