BAYESIAN RANK AGGREGATION FOR VISUAL PLACE RECOGNITION

Frederik Warburg frwa, Martin Skovsbøl Friis s153684, Søren Møller Christensen s153571

Report supplementing the attached project notebook

1. INTRODUCTION

Visual place recognition (VPR) is an essential capability for the long term operation of autonomous robots. It consists of finding an image in a database from the same geographical location as a query image.

VPR is typically solved by using a deep convolutional network to create an embedding for each image. By sorting the distances between the query embedding and the embeddings of all the images in a database, it is possible to get a rank estimate of the most similar images. However, currently, there exist no work on uncertainty quantification (UQ) for these rank estimates.

Currently, the most scalable method for UQ in deep learning is via an ensemble of trained models [1] or its variant MC dropout [2]. In these approaches, an ensemble of model predictions are aggregated to produce a mean prediction with a variance. This is trivial for regression and classification tasks, however, for retrieval tasks, such as place recognition, this aggregation is not trivial.

Research Question: In this project, we investigate a novel, fully bayesian rank aggregation method to create a consensus ranking with associate UQ. We evaluate this method on synthetic data as well as real data.

2. MODELLING SETUP

In this section, we formalize the modeling setup. We have N base rankers that each rank the closest n entries in the database. We define the base-ranker rankings as R_j, where $j = 1 \ldots N$. We assume that the probability density of an observed ranking follows an exponential decreasing function. This assumption is called a Mallows model and is more formally defined as $P(\mathbf{r}|\alpha, \rho) = 1/Z_n(\alpha, \rho) \exp(-\alpha/n \cdot d(\mathbf{r}, \rho))$, where $\mathbf{r}$ is observed ranking, ρ is a center location parameter, α is a scale parameter and $Z_n(\alpha, \rho)$ is normalization term.

Likelihood: Similarly to [3], we model the rank aggregation problem as the joined probability of N Mallows models, allowing us to exploit all available information. Thus, the likelihood function can be defined as

$$P\left(\mathbf{R}_1 ... \mathbf{R}_N | \alpha, \boldsymbol{\rho}\right) = \frac{1}{Z_n(\alpha, \boldsymbol{\rho})^N} \exp\left\{-\frac{\alpha}{n} \sum_{j=1}^{N} d\left(\mathbf{R}_j, \boldsymbol{\rho}\right)\right\}$$

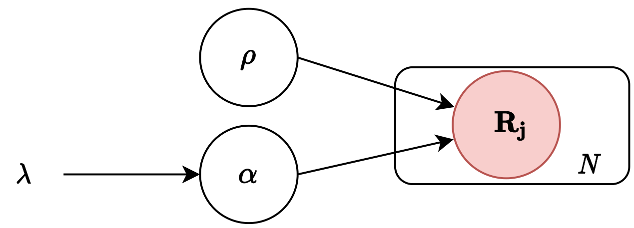

Fig. 1. Probabilistic Graphical Model (PGM) for Bayesian Rank Aggregation.

Thus, the goal of this project is to infer ρ and α. ρ will be the consensus ranking of the N base-rankers, and α will be a scale parameter that describe the degree of agreement between the base-rankers. We show the probabilistic graphical model (PGM) in Figure 1.

Priors: For the scale parameter α, a truncated exponential is chosen with density $\pi(\alpha|\lambda) = \lambda e^{-\lambda\alpha} 1_{[0, \alpha_{max}]}(\alpha)/(1 - e^{-\lambda\alpha_{max}})$.

The prior of ρ is choosen to be a uniform prior over the space of all possible n-dimensional rankings P_n for which $\mathbf{R}_j, \rho \in P_n$.

Posterior: The joined posterior is found by applying Bayes' theorem

$$P\left(\boldsymbol{\rho}, \alpha | \mathbf{R}_1, \ldots, \mathbf{R}_N\right) \propto \frac{\pi(\boldsymbol{\rho})\pi(\alpha)}{Z_n(\alpha)^N} \exp\left\{-\frac{\alpha}{n} \sum_{j=1}^{N} d\left(\mathbf{R}_j, \boldsymbol{\rho}\right)\right\}$$

The consensus ranking is found by marginalizing the joint posterior i.e. integrating the joint posterior over all possible values of α. We will do this in order to find Maximum A Priori (MAP) consensus rank.

3. INFERENCE

The posterior distribution consists of discrete ranks ρ_i and a continuous scale parameter α. Discrete variables are notoriously difficult to deal with in probabilistic frameworks such as Pyro, which relies on gradient information in the approximate inference. When doing inference, Pyro tries to marginalize out the discrete variables, which is troublesome. First it is practically infeasible to marginalize out the discrete variable

ρ_i as it is an $n!$ large space. Second, when marginalizing out the variable, we don't obtain posterior samples, which is one of the main goals of the project. So instead of using gradient based inference in Pyro, we took a similar approach as in [3]. We decided to implement an iterative Metropolis-Hasting sampling method with a leap-&-shift step.

The general Metropolis-Hasting algorithm consists of three steps: (1) generate a proposal sample; (2) calculate an acceptance probability; (3) accept or reject the proposed sample.

In this project we wish to find the posterior of both α and ρ. However, in order to generate a proposal ρ' we need to fix α and similarly, we need to fix ρ when generating a proposal α'. Therefore, we divide the Metropolis-Hasting sampling into two steps: (1) fix α and make inference on ρ; (2) fix ρ and make inference on α. We iterate between these two steps as shown in the pseudo-code in Algorithm 1.

Algorithm 1: MCMC algorithm using L&S

Randomly Initialize ρ_0 and α_0
for $m = 1 \dots M$ **do**
  $\rho' \sim$ L&S(ρ_{m-1}, L)
  $u \sim \mathcal{U}(0, 1)$
  Compute ρ acceptance ratio
  if $u < ratio_\rho$ **then** $\rho_m \leftarrow \rho'$
  else $\rho_m \leftarrow \rho_{m-1}$
  if(m mod α_{jump}) $==$ 0 **then** {
    $\alpha' \sim log\mathcal{N}(\alpha_{m-1}, \sigma_\alpha^2)$ $u \sim \mathcal{U}(0, 1)$
    Compute α acceptance Ratio
    if $u < ratio_\alpha$ **then** $\alpha_m \leftarrow \alpha'$
    else $\alpha_m \leftarrow \alpha_{m-1}$
  }
end

Leap-&-Shift (L&S) is a algorithm for generating a proposal ρ' based on the previous ρ_{m-1}. The intuition of L&S is to flip two entries in the ranking, where it is more likely to flip two neighboring entries than two distant entries. We will elaborate more on this, and how the acceptance ratios can be calculated, in the associated notebook.

4. DATA

We experimented with two datasets: A synthetic one that we generate with ancestral sampling, and a real one generated from the Mapillary Street-Level Sequences (MSLS) [4].

Synthetic The synthetic data is generate with ancestral sampling. We use the following generative story to create the dataset:

1. Draw the scale parameter $\alpha \sim$ TruncatedExponential(λ)
2. Draw the latent consensus ranking $\boldsymbol{\rho} \sim \mathcal{U}(1, n!)$
3. For each model $j \in \{1, \dots, N\}$
   (a) Draw ranking list $\boldsymbol{R}_j \sim \mathcal{L}(\boldsymbol{\rho}, \alpha)$

Mapillary Street-Level Sequences (MSLS) MSLS [4] is the largest, public available dataset for place recognition. In this project we use a subset of 19.195 images. We trained 16 GeM [5] with an AlexNet backbone [6]. After convergence, we calculated the top-100 rankings for the 498 query images in the test set, thus for each of the 16 models we end out with data tensor that is $498 \times 16 \times 100$.

5. RESULTS

Due to space constraints, we only show results from real data in this section. In Figure 2 and 3 we show the posterior for ρ and α for two query images. Here we have used 4 chains with 5000 samples and a 1000 burn-in period for the MCMC algorithm.

Via this bayesian rank aggregation method, we infer the posterior estimates for each entry being at each position. In Figure 2 we show an heatmap over entries belong at a position for two query images. The left heatmap shows two clusters of high intensity. These are artifacts of our completion of the ranking lists, and means that the entries group in the upper left corners are likely to be at the first positions.

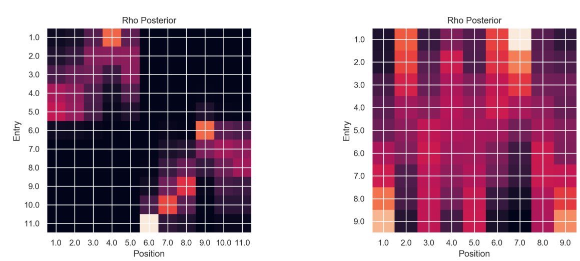

Fig. 2. Posterior for ρ for two different query images.

Figure 3 shows the posterior estimates for α. We see that α is significantly larger in the plot to the left, meaning that the base-rankers are more disagreeing for the first query image.

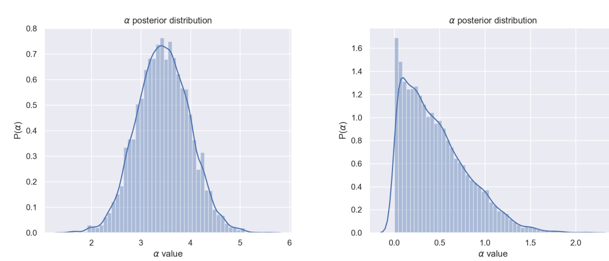

Fig. 3. Posterior for α for two different query images.

6. CONCLUSION

We implemented a bayesian rank aggregation method. To the best of our knowledge, this is the first method to evaluate uncertainty estimate for visual place recognition.

7. REFERENCES

[1] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," 2016.

[2] Yarin Gal and Zoubin Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*. 2016, ICML'16, p. 1050–1059, JMLR.org.

[3] Valeria Vitelli, Øystein Sørensen, Marta Crispino, Arnoldo Frigessi, and Elja Arjas, "Probabilistic preference learning with the mallows rank model," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 5796–5844, Jan. 2017.

[4] Frederik Warburg, Søren Hauberg, Manuel López-Antequera, Pau Gargallo, Yubin Kuang, and Javier Civera, "Mapillary street-level sequences: A dataset for lifelong place recognition," in *Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] Filip Radenovic, Giorgos Tolias, and Ondrej Chum, "Fine-tuning CNN image retrieval with no human annotation," *CoRR*, vol. abs/1711.02512, 2017.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.