Decision making in 2-arm bandit problem

Hannah Chen hechen@eng.ucsd.edu
Sonali Rahagude srahagud@ucsd.edu
Sneha Venkatesh Yelimeli svyelime@eng.ucsd.edu
Chun Fan c9fan@ucsd.edu
Yashodhan Karandikar ykarandi@eng.ucsd.edu

Abstract

1 Introduction

1.1 Motivation

2-arm bandit problems, markov decision process

2 Background

2.1 Related Work

Shuannan's work

3 Design

3.1 Formulation of MDP

states definition, recursive formulation, decision rule

3.2 Environments

3.3 Generation of Optimal data

3.3.1 Algorithm

3.4 Inference of Parameters for Heuristics

For inferring parameters for different heuristic methods, we use the optimal data and do a grid search.

3.4.1 Inference from Optimal Data

3.4.2 Inference from Human Data

3.4.3 Common Algorithm Input : (alpha, beta, decisions, rewards), output: (parameter value)

Algorithm 1 LDA Generative process with collapsed Gibbs Sampling

Input: words $\boldsymbol{w} \in$ documents $\boldsymbol{d} \in [1, D]$

randomly initialize $\boldsymbol{z}$ and increment counters
for iteration $i \in [1, epoch]$ **do**
  for document $d \in [1, D]$ **do**
    for word $\in [1, N_d]$ **do**
      $topic \leftarrow z[word]$
      decrement counters according to document d, $topic$ and $word$
      for $k \in [1, K]$ **do**
        calculate $p(z = k|.)$ using Gibbs equation
      end for
      $newTopic \leftarrow$ sample from $p(z|.)$
      $z[word] \leftarrow newTopic$
      decrement counters according to document d, $newTopic$ and $word$
    end for
  end for
end for

4 Lessons learnt

5 Results

5.1 Parameters by inference from optimal data

add table for decision vectors and reward vectors for different values of alpha, beta

5.2 Parameters by inference from human data

5.3 Match Percentage

5.3.1 Comparison with optimal data

a simple comparison of match percentage of different heuristics with optimal data

5.3.2 Comparison with human data

do we want to include comparison with human data?

6 Conclusions

A Appendix

References

[1] Charles Elkan. Clustering documents with an exponential-family approximation of the dirichlet compound multinomial distribution. In *Proceedings of the 23rd international conference on Machine learning*, pages 289–296. ACM, 2006.

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] Lindsay I Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[4] Gregor Heinrich, Jörg Kindermann, Codrina Lauth, Gerhard Paaß, and Javier Sanchez-Monzon. Investigating word correlation at different scopes—a latent concept approach. In *Workshop Lexical Ontology Learning at Int. Conf. Mach. Learning*, 2005.

[5] Marina Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[6] Gregor Heinrich. Parameter estimation for text analysis. Technical report, 2004.