1. Introduction
2. Problem Definition
3. Literature Survey
4. Approach
   1. Classification
   2. Poisson PMF

Apart from classic classification approach, we also developed another approach, i.e. Poisson probabilistic matrix factorization (PPMF). With PPMF, we can obtain the result of a season at a time.

* + 1. From MF to PPMF

We have learned the matrix factorization (MF) in class: Given a matrix , find a matrix and a matrix such that

where is usually smaller than and , to obtain some kind of “similarity” between the rows and columns of .

To further interpret the data via the concept of probability, probabilistic matrix factorization (PMF) is developed. This method incorporates Bayesian inference by assuming the distribution of the randomness shown by :

where is the probability density function of the Gaussian distribution with mean and variance , and is the indicator of whether entry is presence in . The regularization of are achieved by their priors:

However, the Gaussian distribution model is more suitable for scores whose expectation is far from 0, e.g. NBA. For soccer, the domain of entries of is close to zero and discrete. It is obvious not suitable for Gaussian model.

Thus, we developed a model based on Poisson distribution – the Poisson probabilistic matrix factorization (PPMF).

* + 1. Derivation of PPMF

The likelihood of PPMF is given by:

with the same priors as (3)(4)(就是上两个式子，我就不排了). The log of posterior distribution should be:

where , are a constants that do not depend on the parameters .

Maximizing the log­posterior over with hyperparameters (, ) fixed is equivalent to maximizing the a objective function with quadratic regularization terms:

where and . We can choose values for and by using the soft weight­sharing methods (Steven J. Nowlan and Geoffrey E. Hinton.), but for simplicity, point estimates obtained from the data is used. Minimizing the objective function gives a local minimum, which is a maximum a posteriori (MAP) estimate. PyMC3 is used to find the MAP estimate with Powell optimization.

1. Results
   1. Classification
   2. PPMF
2. Conclusions and Future Work

PPMF can achieve an accuracy of over 40%. Although the accuracy difference between PPMF and PMF is not significant, we believe PPMF is a more suitable model for soccer prediction. For now, only the game goals is used for prediction. In the future, more features of the game should be incorporated.

Reference