

基于贝叶斯理论的万智牌卡牌推荐算法

杨耀飞 李业丽

(北京印刷学院信息工程学院 北京 102600)

摘要 万智牌是一个历史悠久的桌面游戏,因其逻辑复杂且卡牌众多,可以组成大量套牌^[1]。使用卡牌的逻辑进行穷举来推荐卡牌不仅难以实现,而且算法时间复杂度是一个 NP 问题。基于贝叶斯理论的万智牌推荐算法主要利用用户的套牌作为原始数据进行计算得到推荐矩阵,用以替换基于逻辑的推荐算法的逻辑部分,避开了基于逻辑的推荐算法的 NP 问题,而且推荐的准确率也随着用户套牌的增加而增加。

关键词 万智牌,推荐算法,协方差矩阵,相关系数矩阵,机器学习

中图分类号 TP181 文献标识码 A

Magic Cards Recommendation Algorithm Based on Bayesian Theory

YANG Yao-fei LI Ye-li

(College of Information Engineering, Beijing Institute of Graphic Communication, Beijing 102600, China)

Abstract Magic is a long history table games, and can be composed of countless decks because of its complexity in logic and numerous cards. To recommend cards based on logic is difficult to achieve, and the time complexity of the algorithm based on logic is non-deterministic polynomial. Magic recommendation algorithm based on bayesian theory mainly uses the user's decks as the raw data to calculate a recommendation matrix to replace the logic portion of the logic-based recommendation algorithm. It avoids the logic-based recommendation algorithm NP problem, and the recommended rate is increasing with the user accuracy decks increasing.

Keywords Magic card, Recommendation algorithm, Correlationmatrix, Correlation matrix, Machine learning

1 前言

万智牌(Magic The Gathering)是世界上第一个集换式卡牌对战游戏,也是世界性的热门卡牌对战游戏^[1]。

万智牌数量庞大,且每张卡牌的效果也很难用逻辑和多维向量的权值来表达。因每套牌由 60 张卡牌组成,按照一万张卡牌来计算,则可以组成 C_{10000}^{60} 种套牌。所以用穷举加逻辑分析每张牌的效果的方法来对这个系统做卡牌的推荐效率低下,再加上出牌顺序和随机抽牌等规则的限制,使得卡牌推荐系统的算法复杂度为 NP 问题。

本文介绍的基于贝叶斯理论的卡牌推荐算法利用已有的套牌来计算条件概率,并利用推荐卡牌和剩余卡牌之间的条件概率作为推荐基础数据,计算得到与推荐卡牌关联较大的卡牌进行推荐。本算法利用用户操作的数据进行有限的计算,这样也就使基于贝叶斯理论的推荐算法解决了基于逻辑穷举的算法难以实现和计算复杂的问题,且随着用户操作数据的增加,推荐的准确率也随之增加^[2]。

2 基于贝叶斯理论的推荐算法描述

2.1 基于贝叶斯理论的推荐算法的条件概率的计算

因本算法计算条件概率利用的是用户已经存在的套牌数据,每套套牌中卡牌的数量为 60 且可以有重复的卡牌,所以可将每个套牌看成一个集合且其中元素为卡牌和卡牌数量的

集合。

本算法将这个集合变换成一张无向加权图,并导出其邻接矩阵。这样得到的矩阵的每个元的值除以整行的和便可以作为行列数值所对应的卡牌之间的条件概率。因整行每个元都需要除以整行的和,所以不进行这一步,直接用邻接矩阵作为基础条件概率数据即可^[3,4,9]。

2.2 基于贝叶斯理论的推荐算法的条件概率计算的具体描述

(1) 设卡牌牌池为集合 V , 数量为 n , 套牌集合为 S 且套牌元素 s 为集合 $\{n, v\}$, 其中 v 为集合 V 的某张卡牌 n 为卡牌数量。设图 $G(V, E)$, 其中 $E\{p\{v_i, v_j\}, n\}$ 是图 G 的边集合, p 是连接两顶点 v_i, v_j 的边, n 为 E 的权值。

a) 取一套万智套牌 S , 取出其中元素 s_j , 然后遍历 S 。

b) 设当前遍历元素为 s_i , 若在图 $G(V, E)$ 的集合 E 中, 存在元素 $e_k\{p_k, n_k\}$, 其中 p_k 与当前遍历到的元素 s_j 中的 v_j 和 s_i 中的 v_i 组成的 $p\{v_i, v_j\}$ 相等, 则集合 E 中的元素 e_k 中的元素 $n_k = n_k + n_i \times n_j$, 其中 n_i, n_j 为 v_i, v_j 中的元素 n 。

c) 若在图 $G(V, E)$ 的集合 E 中, 不存在元素 $e_k\{p_k, n_k\}$ 与当前遍历到的元素 s_j 中的 v_j 与 s_i 中 v_i 组成的 $p\{v_i, v_j\}$ 相等。则添加 e_k 到 E 中, 其中 $p_k = p, n_k = n_i \times n_j, n_i, n_j$ 为 v_i, v_j 中的元素 n 。

d) 然后去掉 s_i , 重复过程直到 $S = \emptyset$ 。

e) 设有套牌 m 套, 重复进行上述过程, 录入其他套牌。

(2) 根据上述生成的图 G , 获得其邻接矩阵 M , 则根据贝

叶斯定理可得: $p(v_i | v_j) = \frac{M_{[i,j]}}{\sum_{k=1}^n M_{[i,k]}}$

2.3 基于贝叶斯理论的推荐算法的相关性计算

考虑万智牌中存在的某类卡牌,它存在于多数套牌中但对于整套卡牌的贡献值并不高。例如基本地,在以某种思路构建套牌的时候某类基本地都会出现,这样它与以这种思路构建的所有的卡牌所对应的条件概率都很高,但其实这类卡牌不应该被推荐,而应该推荐的是在某些特定卡牌出现的时候条件概率高的卡牌,也就是说应该推荐的是卡牌之间相关性大的卡牌。

根据上述问题,我们对原始邻接矩阵进行处理,得到其协方差矩阵,并进一步得到其相关系数矩阵,并用相关系数矩阵进行卡牌推荐,这样便可去掉上文讨论的不合理的推荐,使推荐更为准确^[5-7]。

2.3.1 相关性计算的预处理

因为原始邻接矩阵的对角线的元都是零,不符合条件概率,所以将它们重置为整行的和。这样处理之后,整个矩阵的对角线上的数据便可以作为自己对自身的条件概率,使得矩阵的每个元素都符合条件概率^[8,10]。

根据上述方法,对于从图 G 中得到的邻接矩阵 M,做如下处理:

$$M_{[i,i]} = \sum_{j=1}^n M_{[i,j]} \quad (1)$$

2.3.2 协方差矩阵和相关系数矩阵的计算

按照式(2)所定义的协方差定义,设 μ_i 为第 i 行的平均值,得协方差矩阵的定义如式(3)所示^[4]。

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (2)$$

$$\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \quad (3)$$

则协方差计算公式如式(4)所示。

$$\text{cov}(M) = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \\ \mu_n \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \\ \mu_n \end{pmatrix}^T \quad (4)$$

按照式(5)所定义的相关系数定义,得到相关系数矩阵的定义及计算公式如式(6)、(7)所示。其中 $\sigma(Y)$ 为 Y 的标准差, M_{ij} 为矩阵 M 的第 i 行、第 j 列的元素。

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad (5)$$

$$\text{cor}(M)_{ij} = \frac{M_{ij}}{\sqrt{M_{ii}M_{jj}}} \quad (6)$$

$$\text{cor}(M) = \begin{pmatrix} (\sqrt{M_{11}})^{-1} & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \times M \times & 0 \\ 0 & 0 & (M_{nn})^{-1} & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & (M_{nn})^{-1} \end{pmatrix} \quad (7)$$

3 算法的效果分析和算法性能分析

3.1 算法相关性计算的算法效果分析

度量相关性计算效果的方法,是手动编写两个协方差矩阵,一个为正常的邻接矩阵,另一个是人为地使某张测试卡牌与其他卡牌都结合成套牌,使用这种人为得出的套牌转换出的邻接矩阵。第一个邻接矩阵测试相关性计算是否影响正常卡牌的推荐,第二个邻接矩阵测试相关性计算是否可以去掉

某些相关性不大,但是在邻接矩阵直接推荐的时候会被推荐的结果。

下面给出两个矩阵:

矩阵一:

0	3	1	1	0	0	0	1
3	0	2	1	0	1	0	1
1	2	0	1	1	1	1	0
1	1	1	0	2	1	0	0
0	0	1	2	0	2	2	0
0	1	1	1	2	1	0	0
0	0	1	0	2	1	0	0
1	1	0	0	0	0	0	0

矩阵二:

0	2	4	3	2	3	2	0
2	0	1	1	0	0	0	0
4	1	0	1	0	1	1	0
3	1	1	0	1	0	0	0
2	0	0	1	0	1	0	0
3	0	1	0	1	0	1	0
2	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0

根据上述矩阵得出如图 1、图 2 所示的相关系数矩阵。

1	-0.258	0.377	-0.0976	-0.313	0.0976	-0.365	0.447
-0.258	1	-0.417	0	-0.135	-0.567	-0.177	0.289
0.377	-0.417	1	0.236	-0.0281	0.236	-0.147	0.602
-0.0976	0	0.236	1	-0.459	0.429	0.802	0.218
-0.313	-0.135	-0.0281	-0.459	1	-0.051	-0.0953	-0.545
0.0976	-0.567	0.236	0.429	-0.051	1	0.535	-0.218
-0.365	-0.177	-0.147	0.802	-0.0953	0.535	1	-0.408
0.447	0.289	0.602	0.218	-0.545	-0.218	-0.408	1

图 1 相关系数矩阵 1

1	-0.134	-0.463	-0.39	-0.267	-0.39	-0.134	-2
-0.134	1	0.722	0.73	0.75	0.73	0.75	-2
-0.463	0.722	1	0.738	0.866	0.738	0.722	-2
-0.39	0.73	0.738	1	0.548	0.867	0.73	-2
-0.267	0.75	0.866	0.548	1	0.548	0.75	-2
-0.39	0.73	0.738	0.867	0.548	1	0.73	-2
-0.134	0.75	0.722	0.73	0.75	0.73	1	-2
-2	-2	-2	-2	-2	-2	-2	-2

图 2 相关系数矩阵 2

根据上述矩阵,分析后得出相关性计算不仅没有改变邻接矩阵的推荐序列,而且去掉了我们认为不应该推荐的卡牌,这说明相关性计算是有效的。

3.2 实验结果分析

用于实验的卡牌的整个万智牌牌池为 1400 张牌,并利用这 1400 张卡牌构筑套牌 30 套,每套将套牌的构筑威胁性、构筑控制性和构筑维持性 3 个套牌属性作为套牌的权值,并分别运用本算法计算得出一个推荐矩阵,并将推荐矩阵导入数据库放在 Web 上做成网页进行显示和推荐,得出某推荐效果如图 3 所示,将推荐结果整理成文字如表 1 所列。



图 3 Web 网站推荐效果

表1 米捷轰击推荐效果

构筑思路	推荐 1	推荐 2	推荐 3	推荐 4	推荐 5
构筑威胁性	波洛斯	火焰大师	圣洁庇护	波洛斯护符	胜利殿堂
构筑控制性	火焰大师	波洛斯	岚惠巨龙	众神之怒	波洛斯护符
构筑维持性	岚惠巨龙	火焰大师	铅直落下	贪欢神谕	波洛斯裁定师

根据表1的推荐,这个推荐结果在3个方面都具有较强的参考价值,威胁型策略排名第一的推荐为相对低费用的进攻生物,便于早期建立优势,第二名是旅法师用于扩大优势持续威胁对手;而控制型策略则首先推荐了旅法师用于改变场面控制对手;对于维持型策略,第一选择是高费用的强大生物,提供了维持到游戏后期的取胜所用的卡牌。

3.3 算法性能分析

为了便于下述的算法分析,设卡牌数量为 n ,套牌数量为 m ,且已知套牌中卡牌数量为60。

3.3.1 算法时间复杂度分析

将一套万智套牌转换成无向加权图所需要的时间复杂度为常数,是添加一条无向加权图的边所需时间的 C_0 倍。又已知万智套牌的数量为 m ,所以将所有 m 套万智套牌转换成无向加权图的时间复杂度为 $O(m)$ 。

因输出的邻接矩阵是一个 $n \times n$ 矩阵,而相关性计算过程中进行了多次矩阵相乘和矩阵的行列间计算,行列间计算的时间复杂度为 $O(n^2)$,矩阵相乘的时间复杂度为 $O(n^3)$,所以相关性计算部分的时间复杂度为 $O(n^2 + n^3) = O(n^3)$ 。

综上所述,整个算法的时间复杂度为 $O(n^3 + m)$ 。

3.3.2 算法的空间复杂度分析

同上所述,我们先对万智套牌转换成无向加权图的过程的空间复杂度进行分析。转换成无向加权图的过程中的空间使用集中在储存无向加权图上,而转换过程中的计算过程所需要的储存空间为常数。储存无向加权图使用了邻接链表,每处理一套万智套牌,最多会增加 C_0 个节点,又因为邻接链表中点的数量为卡牌的数量,所以储存邻接链表所需要的储存空间为 $n + m \times C_0$ 。所以将套牌转换成无向加权图的过程的空间复杂度为 $O(n + m)$ 。

无向加权图输出的是一个邻接矩阵,而在相关性计算过程中也是矩阵的相关计算,计算过程中的空间的使用集中在矩阵的储存上,计算过程中的临时变量所用的空间为常数。所以相关性计算过程中的空间复杂度为 $O(n^3)$ 。

综上所述,整个算法的空间复杂度为 $O(n^3 + m)$ 。

结束语 基于贝叶斯理论的万智牌卡牌推荐算法的核心思想是利用大量的万智套牌进行计算,得出万智牌之间的相

关性,并利用相关性进行推荐。在方法上,本算法利用了无向加权图和邻接矩阵,将套牌中的多对多关系转换成了条件概率矩阵,并对该矩阵进行变换以得出万智牌之间的相关性。

本算法的优点是在卡牌一定的情况下,算法可以在有限时间内完成,且每次推荐的过程不需要运行一次算法,只需要对算法得出的推荐矩阵的被推荐行进行排序便可得出推荐序列,这样可以大大减少推荐过程中运行本算法的计算负担,加快推荐速度。本算法也有不足和需要改进的地方,因为推荐矩阵的计算需要大量套牌作为输入,所以套牌的数量和套牌的质量决定了本算法最终的准确性。除此外,本算法是一个静态的算法,当出现新的万智套牌时不能直接加入推荐矩阵,需要重新计算。这也是本算法需要改进的地方——增加动态添加推荐矩阵的方法。

参考文献

- [1] Wikipedia.org. Magic: The Gathering [EB/OL]. 2014-04-02 [2014-04-03]. http://en.wikipedia.org/wiki/Magic:_The_Gathering
- [2] Wikipedia.org. Recommender System [EB/OL]. 2013-05-05 [2014-04-01]. http://en.wikipedia.org/wiki/Recommender_system
- [3] Stanly B. Lippman Josee Lajoie, Barbara E Moo. C++ Primer 4th Edition. [M]; Addison-Wesley Professional, 2005; 623
- [4] D'Agostini G. Bayesian reasoning in data analysis [M]. World Scientific, 2003
- [5] Wu J, Chen L, Jian H, et al. Composite service recommendation based on Bayes theorem[J]. International Journal of Web Services Research (IJWSR), 2012, 9(2): 69-93
- [6] Lau C W. News Recommendation System Using Logistic Regression and Naive Bayes Classifiers[J]. 2011
- [7] Salehi M, Nakhai Kamalabadi I. A hybrid recommendation approach based on attributes of products using genetic algorithm and naive Bayes classifier[J]. International Journal of Business Information Systems, 2013, 13(4): 381-399
- [8] Lisboa P, Nawaf H, Bhaya W. Improving Recommendation Systems by Modeling the Stability of Implicit Behaviour[C]// The Post Graduate Network Symposium (PGNet2013). Liverpool, UK. 2013
- [9] de Groot H T. iNewsReader: Personal Netnews recommendation using Naive Bayes & Support Vector Machines[D]. Netherlands: University of Groningen, 2011
- [10] Puntheeranurak S, Pitakpaisarnsin P. Time-aware Recommender System Using Naive Bayes Classifier Weighting Technique[C]// 2nd International Symposium on Computer, Communication, Control and Automation. Atlantis Press, 2013; 266-269

(上接第64页)

- [5] Zhao J Y, Li M, Zhang W W, et al. ECG Signal Adaptive Filtering and QRS Complex Detecting Method[C]// 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010). Oct 2010; 869-872
- [6] Dash S, Shelley K H, Silverman D G, et al. Transducer Signals: A Comparative Study of Time-Frequency Methods[J]. IEEE Transaction on Biomedical Engineering, 2010, 57(5): 1099-1106
- [7] Daubechies I. The wavelet transform, time-frequency localization and signal analysis[J]. IEEE Transactions on Information Theory, 1990, 36(5): 961-1005
- [8] Ibaida A, Khalil I. Wavelet-Based ECG Steganography for Pro-

tecting Patient Confidential Information in Point-of-Care Systems[J]. IEEE Transactions on Biomedical Engineering, 2013, 60(12): 3322-3330

- [9] 徐晓刚, 徐冠雷, 王孝通, 等. 经验模式分解(EMD)及其应用[J]. 电子学报, 2009, 37(3): 581-585
- [10] Labate D, Foresta F L, Occhiuto G, et al. Empirical Mode Decomposition vs. Wavelet Decomposition for the Extraction of Respiratory Signal from Single-Channel ECG: A Comparison[J]. IEEE Sensors Journal, 2013, 13(7): 2666-2674
- [11] Khaldi K, Boudraa A O. On signals compression by EMD[J]. Electronics Letters, 2012, 48(21): 1329-1331