Time-Interest Coupled IPTV User Behavior Model

Ya Zhang
Shanghai Jiao Tong University

Joint Work with Weiyuan Chen and Hongyuan Zha

Internet Protocol Television (IPTV)

- IPTV: Deliver television services using the Internet.
- Video on demand (VoD): Users select and watch video on demand.

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Modeling User Behavior in IPTV

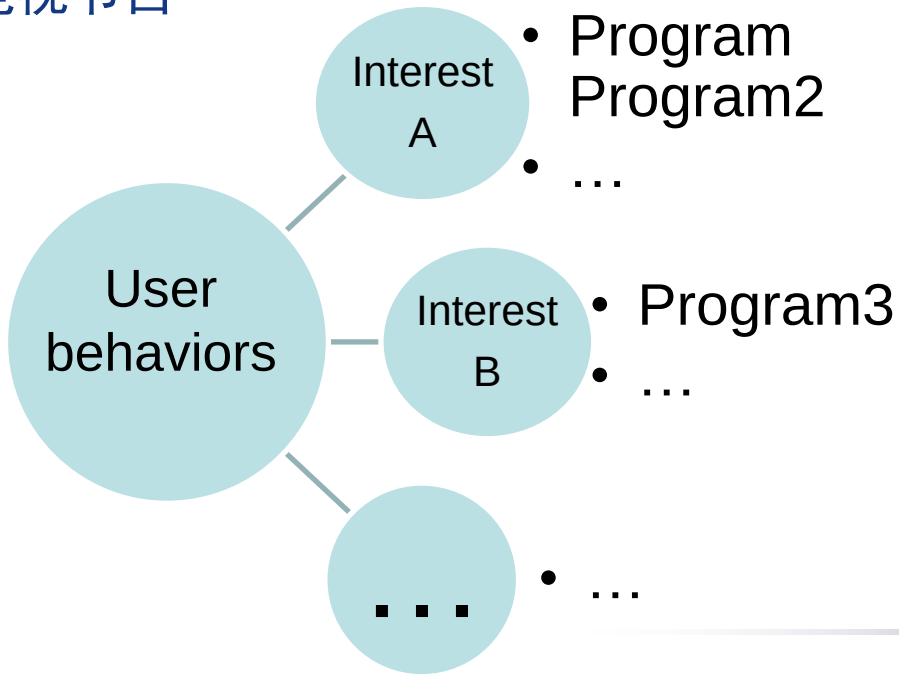
- Characterizing and modeling the user activities in an IPTV network
  - Improve the IPTV system efficiency, e.g. channel switching
  - Recommender systems, e.g. program recommendation, personalized EPG, targeted advertisement

Latent Dirichlet Allocation

- LDA is a generative probabilistic model originally proposed for a corpus. The basic idea is that the documents are represented as random mixtures over latent topics, where a topic is characterized by a distribution over words.
- In the IPTV case, user behaviors are represented as random mixtures over latent interests, where an interest is characterized by a distribution over programs.

LDA-based IPTV User Behavior Model

- 用户的点播行为就可以由以下三步生成：
  - 首先，对每一个IPTV 用户，从一个狄利克雷分布中采样出一个兴趣偏好的分布
  - 然后，对用户的每一次点播行为，从该兴趣偏好分布中采样出一个兴趣偏好
  - 最后，从这个选出的兴趣偏好在电视节目上的多项分布中采样出一个电视节目



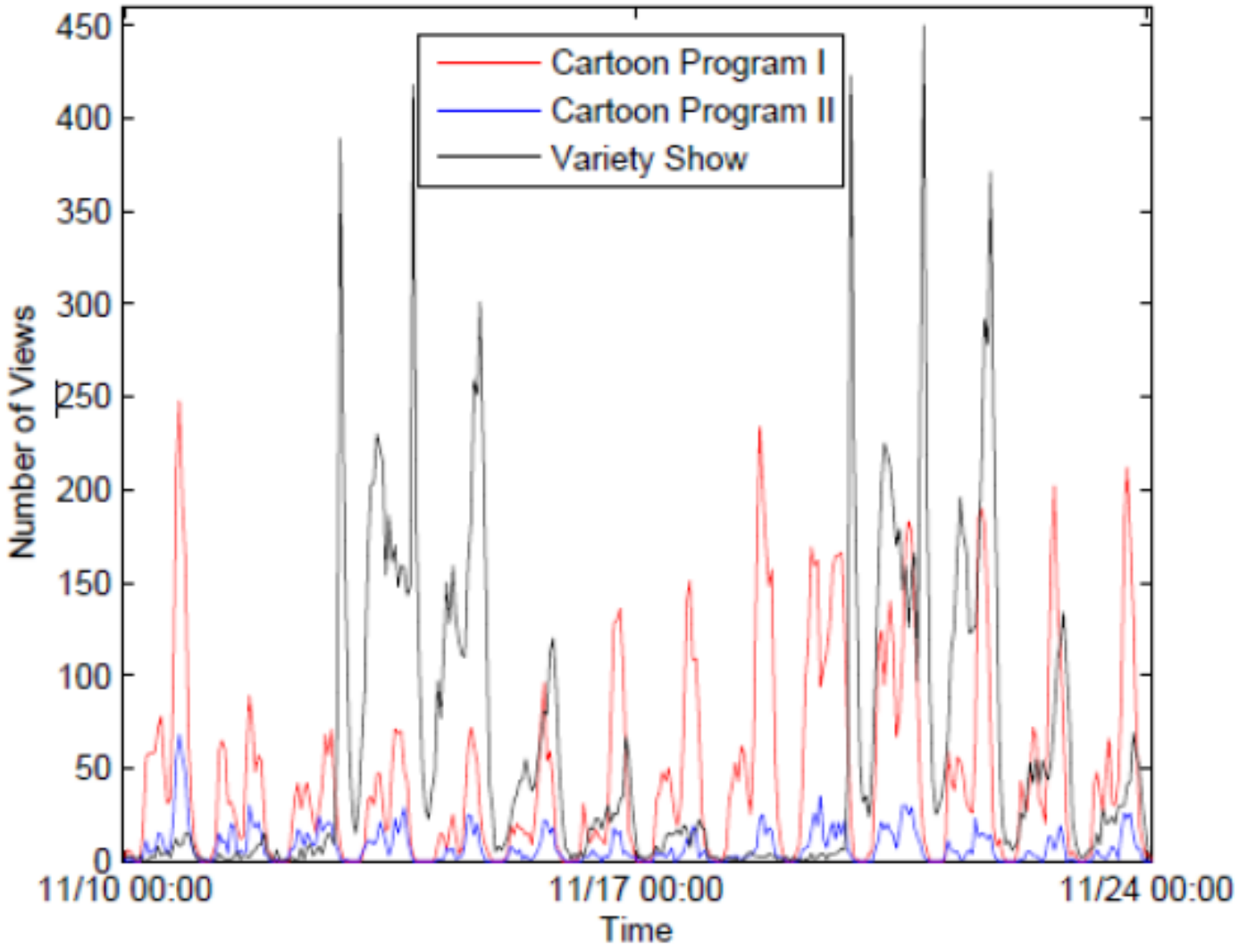
Characteristics of IPTV data

- A family has one or several members.
- Each member can have a diverse set of **interests**.
- The interests of members may be different.
- Each member tends to watch TV at certain time periods every week, resulting in **temporal viewing patterns**.

Challenge: How to tell who is watching in a family?

e.g. Do not want to recommend adult content when kids are watching and vise versa

Temporary Viewing Patterns



Behavior Patterns

- Behavior Pattern: When a family likes to watch what kind of TV shows.

Family ID	196843d1bb	
Starting Time	Program ID	Program Title
2011-12-30 14:35:51	440929	The Black Fox
2011-12-30 15:37:41	442425	The Black Fox
2011-12-30 18:22:33	317986	Tom and Jerry

Temporal viewing patterns

Interests

- Mining common behavior patterns
- Characterizing family structures and users

Coupled LDA(cLDA) Model

- Each family is described by a distribution Θ of interests and temporal patterns.

Temporal Patterns

	17~19PM weekdays	20~22PM weekdays	13~16PM weekdays	13~18PM weekends
Cartoon	0.18	0.01	0.01	0.1
Variety	0.01	0.2	0.15	0.04
War	0.01	0.1	0.01	0.18

Interests

- Each interest is described by a distribution ϕ of TV Programs.

	Tom and Jerry	Doraemon	Talent show	The Voice
Cartoon	0.5	0.4	0.05	0.05
Variety	0.04	0.03	0.52	0.41

- Each temporal pattern is described by a distribution ψ of timestamps.

cLDA-based IPTV User Behavior Model

- 用户的点播行为就可以由以下四步生成：

第一步：选择用户 m 的一种行为模式分布 $\vec{\theta}_m \sim Dir(\alpha)$，其中 $m \in [1, M]$；

第二步：选择兴趣偏好 k 下的一种电视节目分布 $\vec{\phi}_k \sim Dir(\beta)$，其中 $k \in [1, K]$；

第三步：选择时间模式 l 下的一种时间戳分布 $\vec{\psi}_l \sim Dir(\gamma)$，其中 $l \in [1, L]$；

第四步：对于用户 m 的第 n 次点播行为，其中 $n \in [1, N_m]$

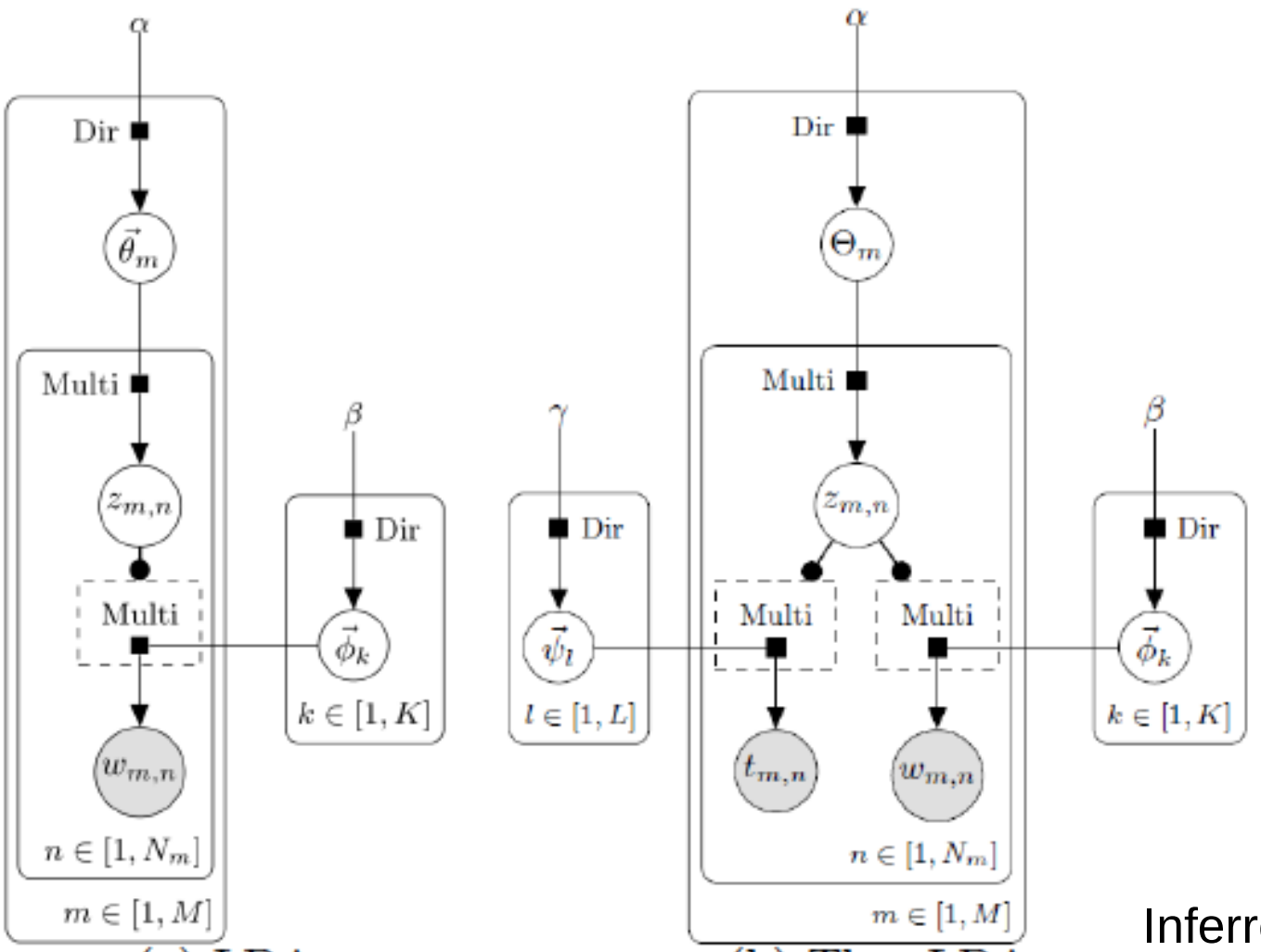
（a）选择该行为的一种行为模式 $z_{m,n} \sim Multi(\vec{\theta}_m)$；

（b）选择该行为的兴趣偏好下的一个电视节目 $w_{m,n} \sim Multi(\vec{\phi}_{z_{m,n,1}})$；

（c）选择该行为的时间模式下的一个时间戳 $t_{m,n} \sim Multi(\vec{\psi}_{z_{m,n,2}})$。

（d）$(w_{m,n}, t_{m,n})$ 二元组即构成了这次点播行为。

Graphical Model



Inferred with Gibbs sampling

LDA vs. cLDA

	Interest	Temporal Pattern	Characterize Family by
LDA	√	×	interests
cLDA	√	√	Interests and temporal patterns

(a) $\vec{\theta}_m$ in LDA model

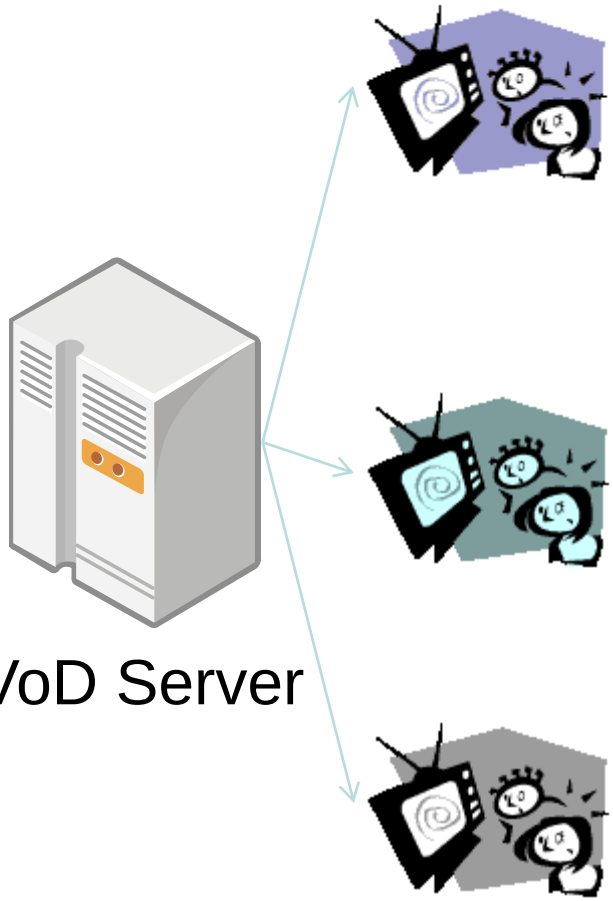
Cartoon	0.3
Variety	0.4
War	0.3

(b) Θ_m in the cLDA Model

	17∼19PM weekdays	20∼22PM weekdays	13∼16PM weekdays	13∼18PM weekends
Cartoon	0.18	0.01	0.01	0.1
Variety	0.01	0.2	0.15	0.04
War	0.01	0.1	0.01	0.18

Dataset

- IPTV log

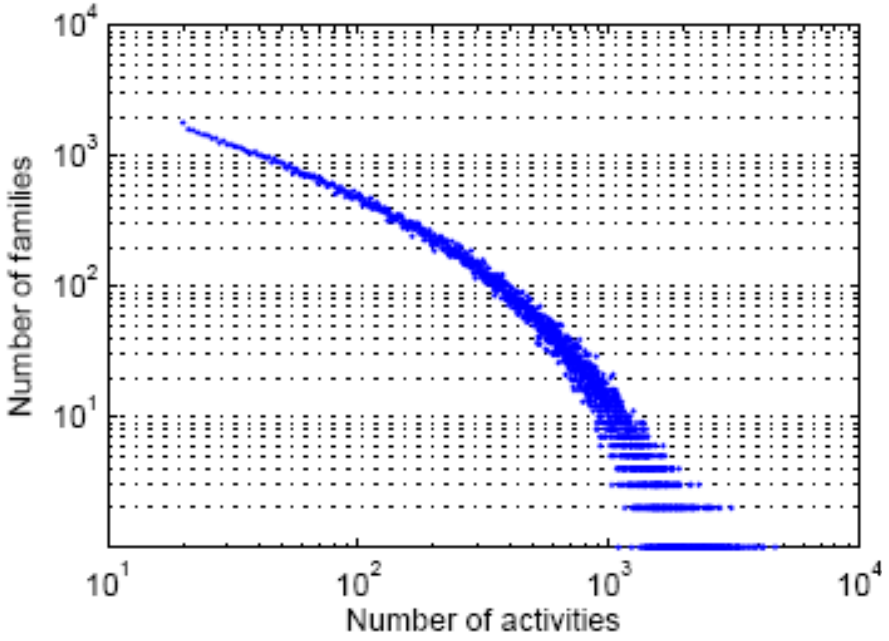


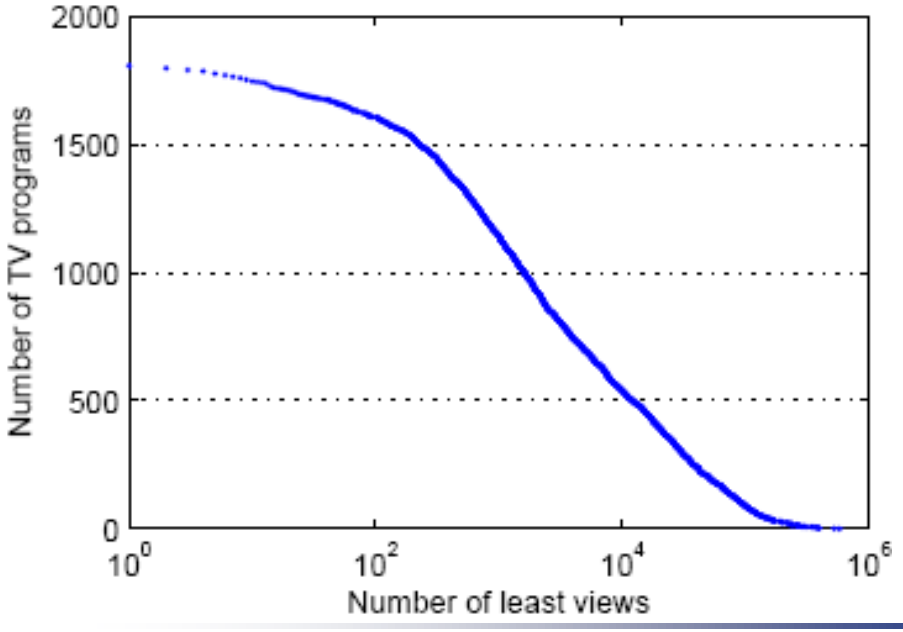
Family ID		196843d1bb
Starting Time	**Program ID**	**Program Title**
2011-12-30 14:35:51	440929	The Black Fox
2011-12-30 15:27:41	442425	The Black Fox
Family ID		**1968470219**
Starting Time	**Program ID**	**Program Title**
2011-12-30 20:55:10	444352	Golden Code
2011-12-30 21:24:33	444109	A Tale of Two Cities

.....

Dataset

- Basic statistics
  - 9 months of log
  - 1805 TV programs
  - 154622 families
  - over 34 million records.





卡通类兴趣偏好（K=50）

兴趣 35	兴趣 39	兴趣 40
盖亚奥特曼	天火传说	斗龙战士
迪迦奥特曼	金甲战士	奇幻龙宝
泰罗奥特曼	果宝特攻	猫眼三姐妹
铠甲勇士	铠甲勇士	霹雳乐翻天
星际恐龙	钢甲卡卡龙	魔角侦探
宇宙英雄奥特曼	恐龙宝贝	小青天司徒公
金甲战士	锋速战警	星际恐龙
锋速战警	火力少年王 4	超级偶像叮当猫
超人	闪电冲线 2	夺宝幸运星
超音战士	火力少年王 2	恐龙宝贝

综艺类兴趣偏好（K=50）

兴趣 8	兴趣 18	兴趣 48
康熙来了	新娱乐在线	非诚勿扰
康熙	百里挑一	弈棋耍大牌
国光帮帮忙	谁能百里挑一	百里挑一
娱乐百分百	家庭演播室	壹周立波秀
大学生了没	回顾	谁能百里挑一
大小爱吃	第三季中国达人秀	时尚中国
快乐大本营	中国达人秀	爱情连连看
天天向上	36.7℃明星听诊会	非诚勿扰精华版
新片推手	相约星期六	晚间新闻
沈春华 life 秀	星光大道	寻宝

韩剧类兴趣偏好（K=50）

兴趣 1	兴趣 12
秘密花园	妻子的诱惑 I
游戏的女王	妻子的诱惑 II
风之画师	妻子的诱惑 III
风格	游戏的女王
女王的条件 III	风格
偷心大圣 PS 男	最爱
天桥风云	阿岘洞夫人 I
初恋	妻子们的战争 I
花美男拉面馆	犀利人妻
巴厘岛的日子	天使的诱惑

年轻的未婚女性　　已婚的中年妇女

LDA vs cLDA：Interests Discovered

- Dominated interests: the interest dominated by less than 10 TV shows. The dominated interests tend to group irrelevant TVs together.
- LDA generates 9 dominated interests, the cLDA generate 5 dominated interests.
- cLDA generates more coherence interests than LDA.

Program Title	Probability
Palace: The Locked Heart Jade	0.7848
Opposite Attraction III	0.0439
Opposite Attraction II	0.0409

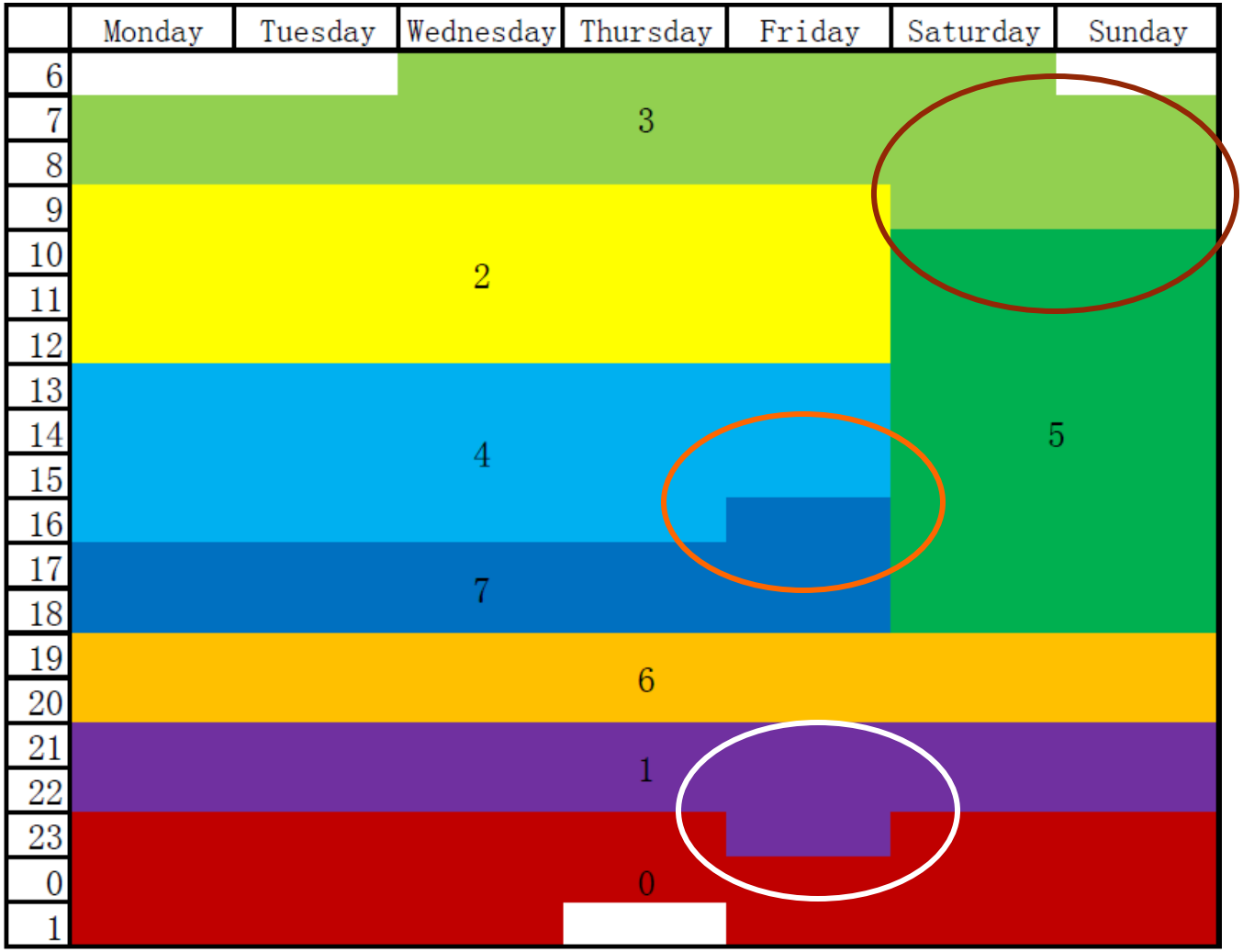
An interest in LDA (Dominated Interest)

Program Title	Probability
Palace: The Locked Heart Jade	0.5728
Schemes of a Beauty	0.3627
Happy Mother-in-law, Pretty Daughter-in-law	0.0492

An interest in cLDA (coherence interest)

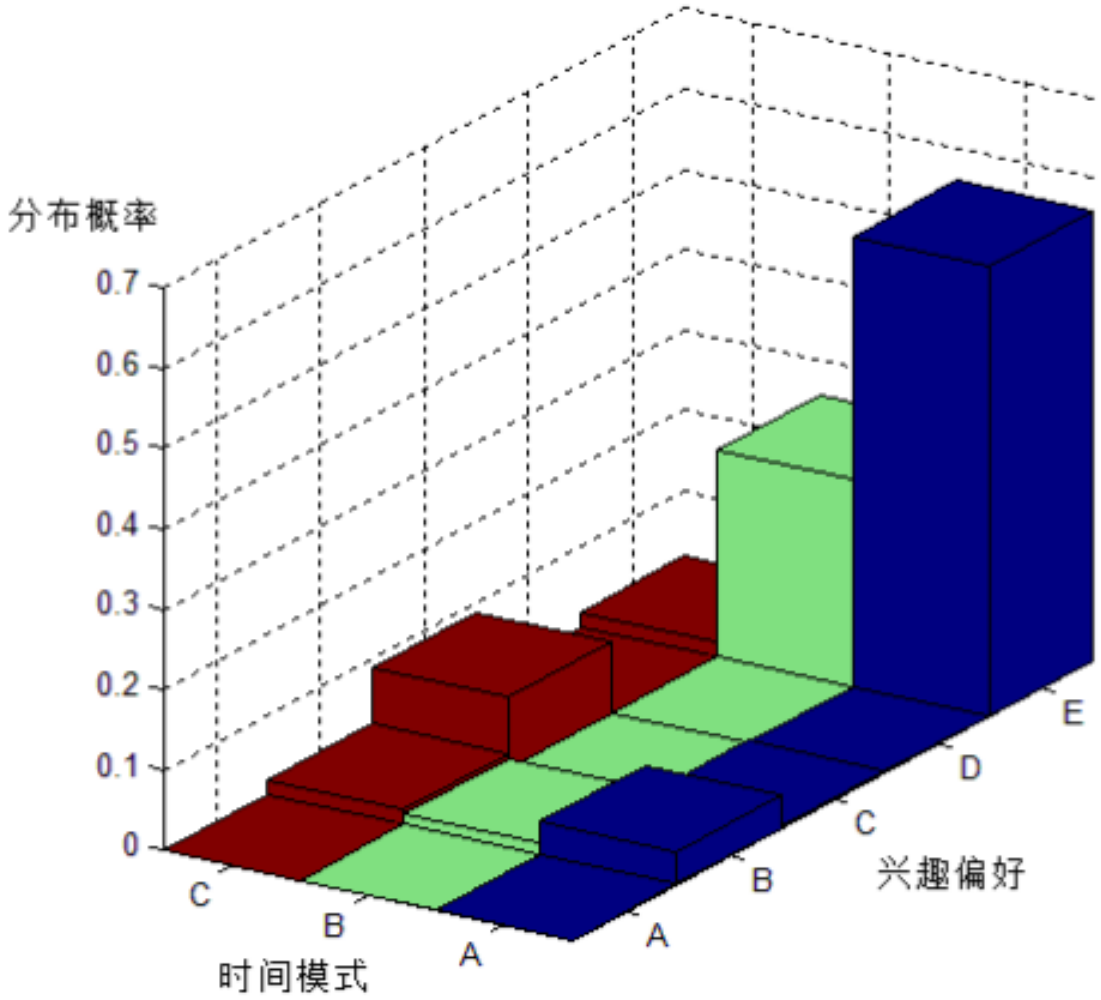
Titles in red are Chinese romance historical fiction shows.
Titles in yellow are Korean modern love show .

Temporal Patterns



L=8

儿童主导型家庭的行为模式分布

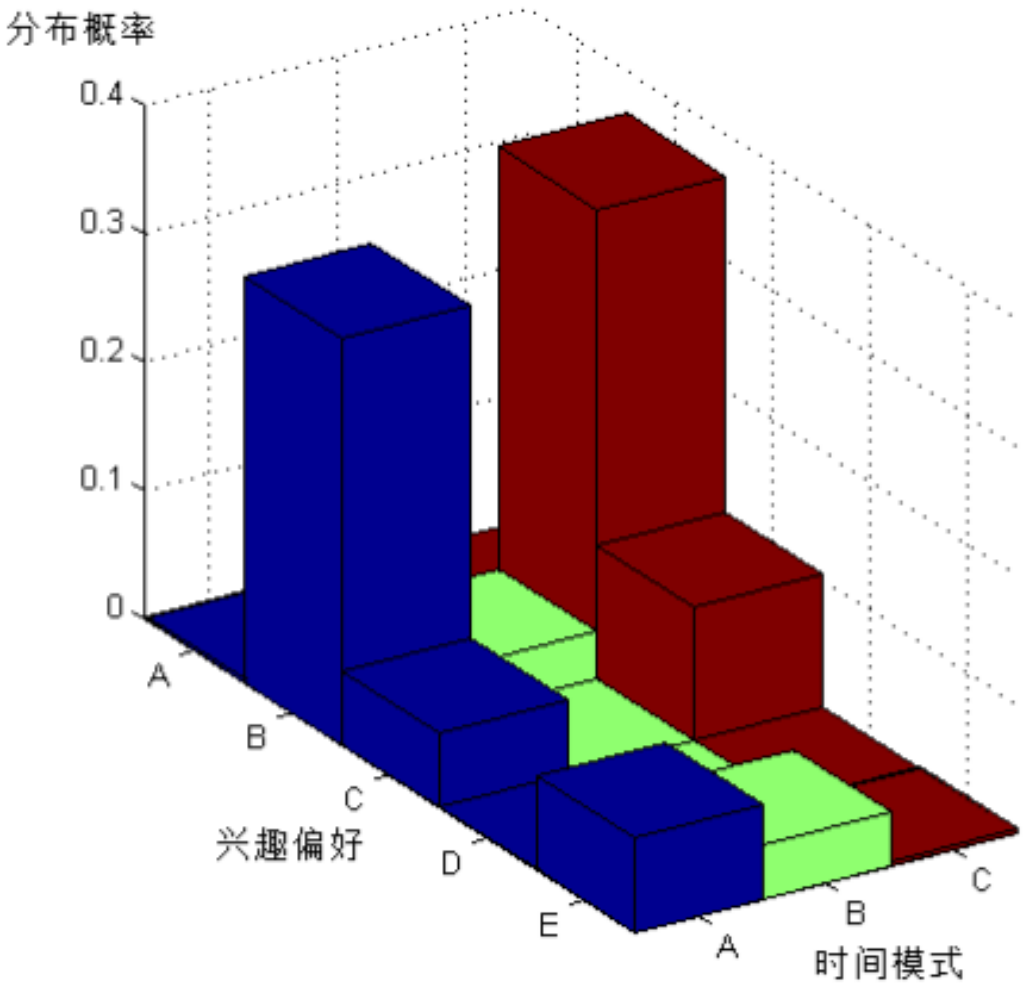


兴趣A
战争、悬疑类电视剧，古装历史剧；
兴趣B
综艺娱乐节目、古装宫廷剧、爱情喜剧；
兴趣C
当代都市家庭剧、抗战剧、社会新闻类栏目；
兴趣D
当代家庭生活剧、青春偶像剧、韩剧；
兴趣E
卡通动漫、古装历史剧。

时间A：周一至周五17至21点，周六周日11点至22点；
时间B：周一至周五7点至17点；
时间C：每天22点至第二天6点。

K=5，L=3

年轻上班族家庭行为模式分布

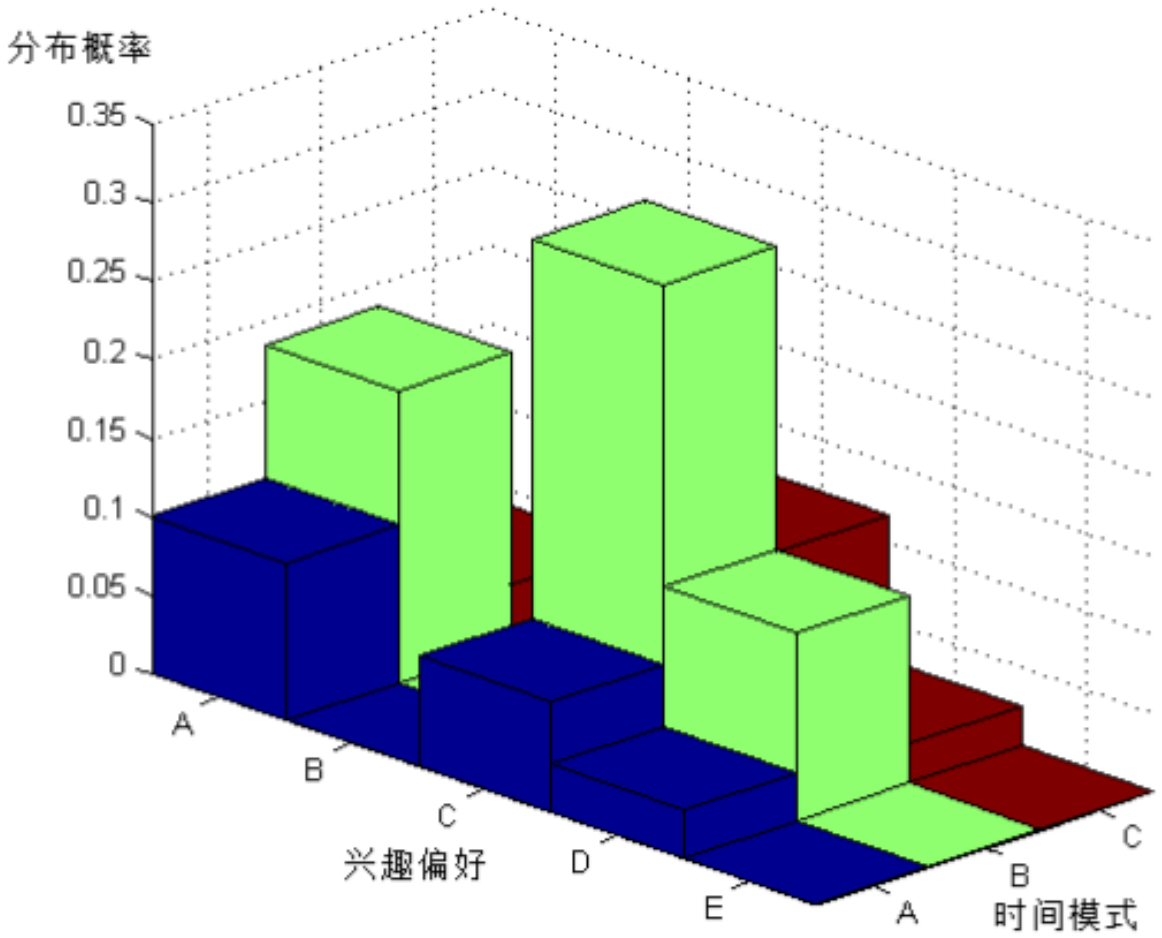


兴趣A
战争、悬疑类电视剧，古装历史剧；
兴趣B
综艺娱乐节目、古装宫廷剧、爱情喜剧；
兴趣C
当代都市家庭剧、抗战剧、社会新闻类栏目；
兴趣D
当代家庭生活剧、青春偶像剧、韩剧；
兴趣E
卡通动漫、古装历史剧。

时间A：周一至周五17至21点，周六周日11点至22点；
时间B：周一至周五7点至17点；
时间C：每天22点至第二天6点。

K=5，L=3

退休夫妻家庭的行为模式分布

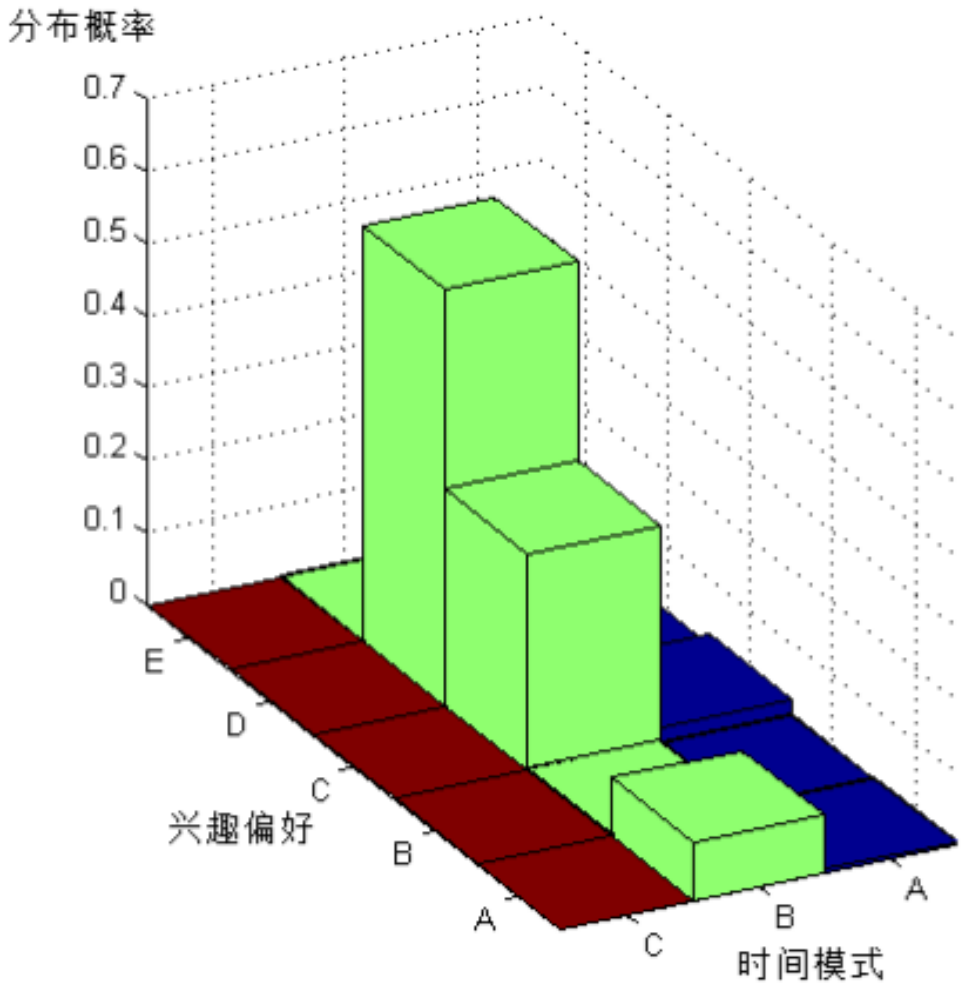


兴趣A
战争、悬疑类电视剧，古装历史剧；
兴趣B
综艺娱乐节目、古装宫廷剧、爱情喜剧；
兴趣C
当代都市家庭剧、抗战剧、社会新闻类栏目；
兴趣D
当代家庭生活剧、青春偶像剧、韩剧；
兴趣E
卡通动漫、古装历史剧。

时间A：周一至周五17至21点，周六周日11点至22点；
时间B：周一至周五7点至17点；
时间C：每天22点至第二天6点。

K=5，L=3

家庭主妇型家庭的行为模式分布



兴趣A
战争、悬疑类电视剧，古装历史剧；
兴趣B
综艺娱乐节目、古装宫廷剧、爱情喜剧；
兴趣C
当代都市家庭剧、抗战剧、社会新闻类栏目；
兴趣D
当代家庭生活剧、青春偶像剧、韩剧；
兴趣E
卡通动漫、古装历史剧。

时间A：周一至周五17至21点，周六周日11点至22点；
时间B：周一至周五7点至17点；
时间C：每天22点至第二天6点。

K=5，L=3

Case Analysis

- Examine the behavior patterns generated by cLDA for a family that mainly watches cartoon.

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
6	0	0	0	0	0	1	0
7	5	7	1	0	1	0	0
8	14	25	26	10	20	3	7
9	12	9	29	9	23	5	2
10	13	26	33	21	29	4	1
11	28	33	36	39	47	8	3
12	23	24	27	33	27	6	4
13	1	2	4	1	0	0	1
14	0	0	0	0	1	0	2
15	10	12	9	14	7	1	3
16	19	32	7	18	20	5	4
17	9	7	4	7	4	5	14
18	8	6	1	5	4	12	9
19	3	0	2	0	3	6	0
20	1	0	2	0	1	0	2

Cartoon's Watching Time of the Family

Case Analysis

- The patterns generated by cLDA successfully tells that the family likes to watch cartoons at 9 to 12AM.

Temporal Pattern	0 (11PM~1AM everyday)	2 (9~12AM weekday)	7 (17~18PM weekday)
Cartoon for infant	0.0006	**0.3042**	**0.0530**
Cartoon for child	0.0006	**0.2343**	**0.0388**
Chinese War Show	**0.0271**	0.0084	0.0006

Patterns Distribution Generated by cLDA

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
6	0	0	0	0	0	1	0
7	5	7	1	0	1	0	0
8	14	25	26	10	20	3	7
9	12	9	29	9	23	5	2
10	13	26	33	21	29	4	1
11	28	33	36	39	47	8	3
12	23	24	27	33	27	6	4
13	1	2	4	1	0	0	1
14	0	0	0	0	1	0	2
15	10	12	9	14	7	1	3
16	19	32	7	18	20	5	4
17	9	7	4	7	4	5	14
18	8	6	1	5	4	12	9
19	3	0	2	0	3	6	0
20	1	0	2	0	1	0	2

Cartoon's Watching Time of the Family

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Program Recommendation

- Task: Predict what the TV program the family would choose when they turn on their TV at a certain time.
- Procedure:
  - First train a cLDA model using the fully observed set of families (the test data is excluded).
  - For test families, we are shown all but the activities of the entire last day.
  - We are told when each of the held-out activities occurs.
  - Predict what TV-program the family would have chosen at the given time.

Program Recommendation

- The number of interests for each model ranges from 5 to 50.
- Evaluation Metric: Predictive Perplexity

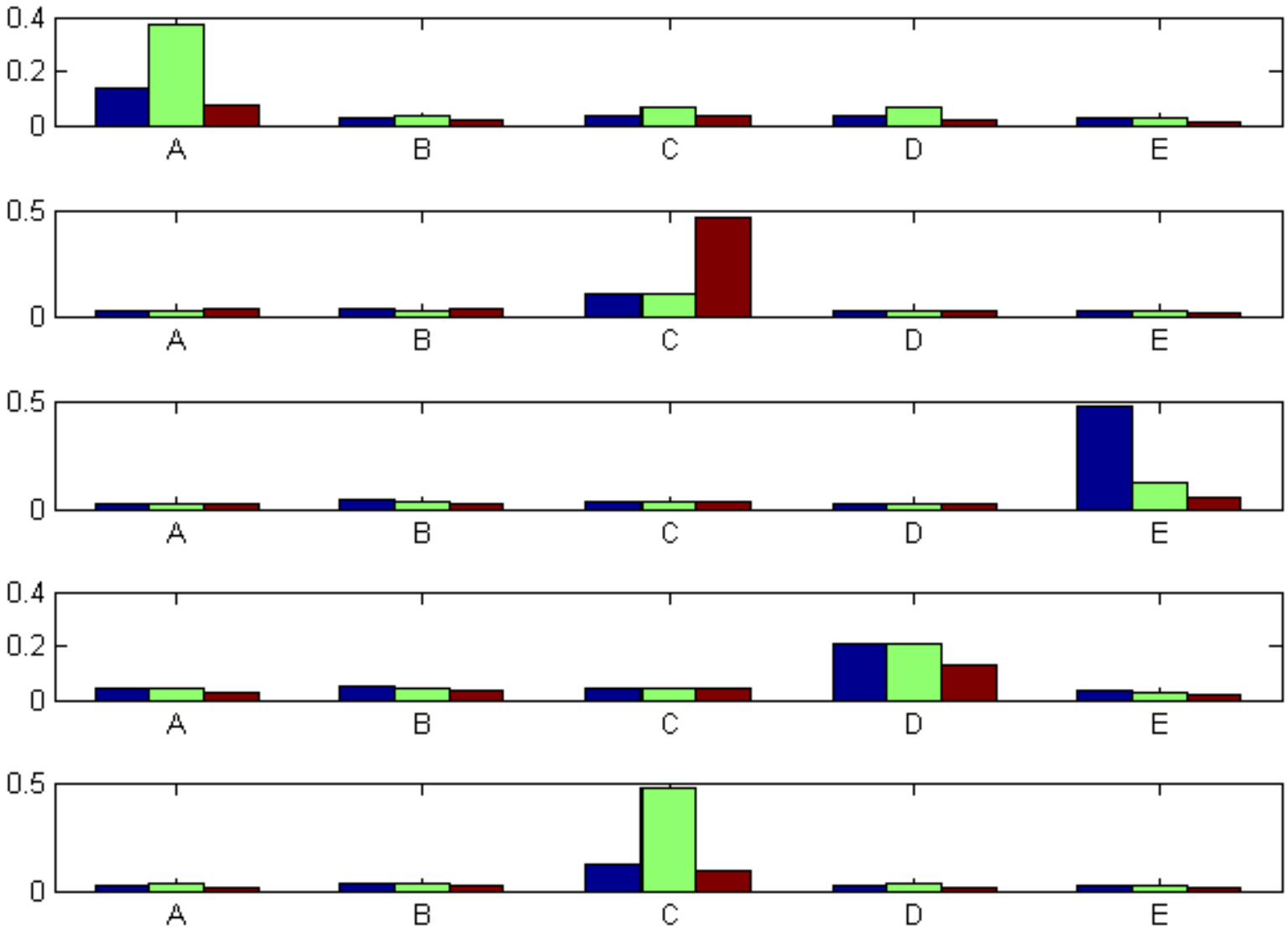
$$predictive - perplexity(D_{test})$$
$$= \exp\{ - \frac{\sum_{m=1}^{M_{test}} \log p(w_{m,N_m} | \vec{w}_{m,1:N_m-1}, \vec{t}_{m,1:N_m})}{M_{test}} \}$$

K	5	10	20	50	100
LDA	458.1	376.5	306.8	227.9	185.5
cLDA	396.6	357.6	276.5	221	179

Predictive-perplexity

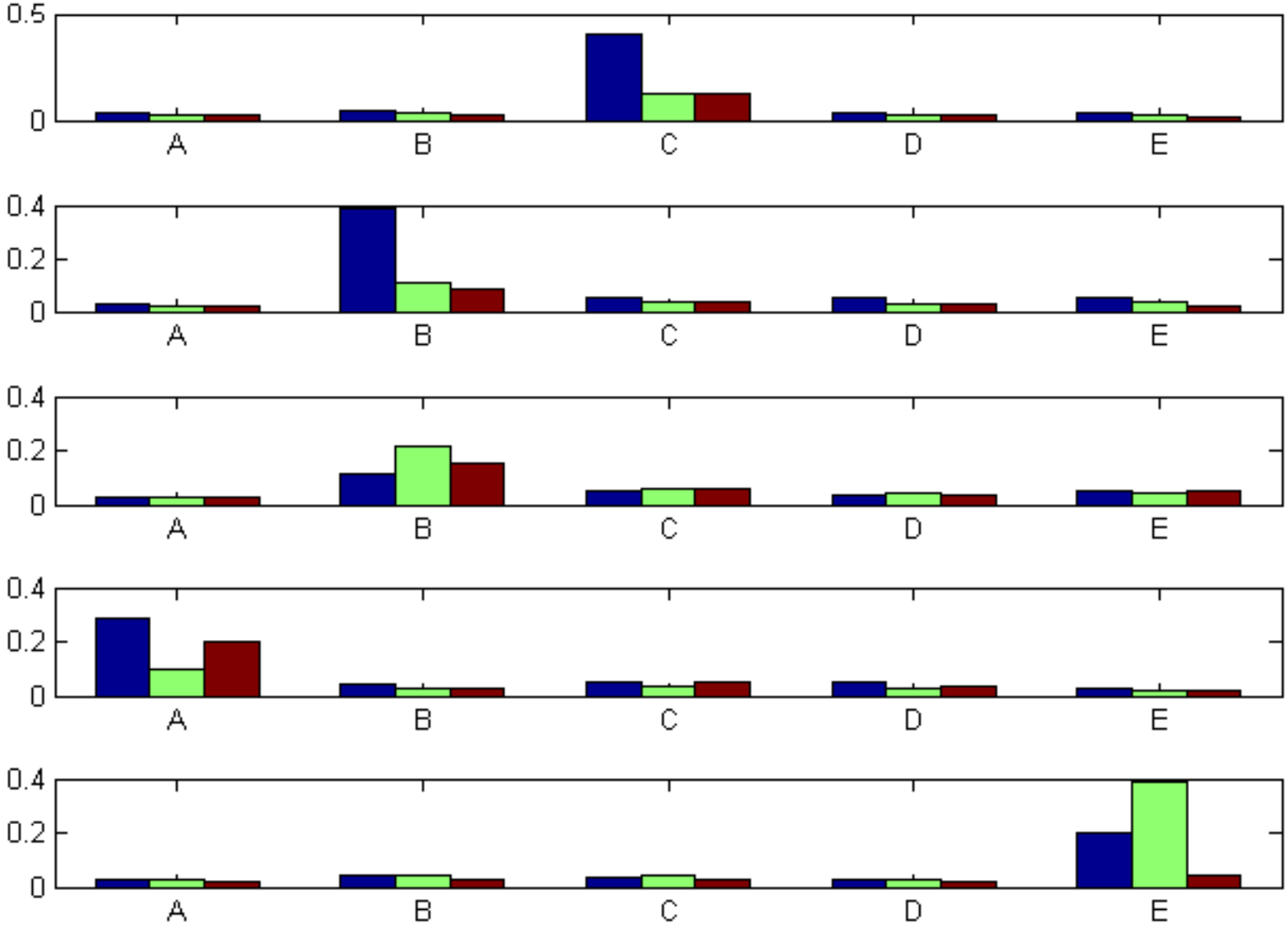
A lower perplexity score indicates a better generalization performance

用户类中心点的行为模式分布



（K=5、L=3、C=10）

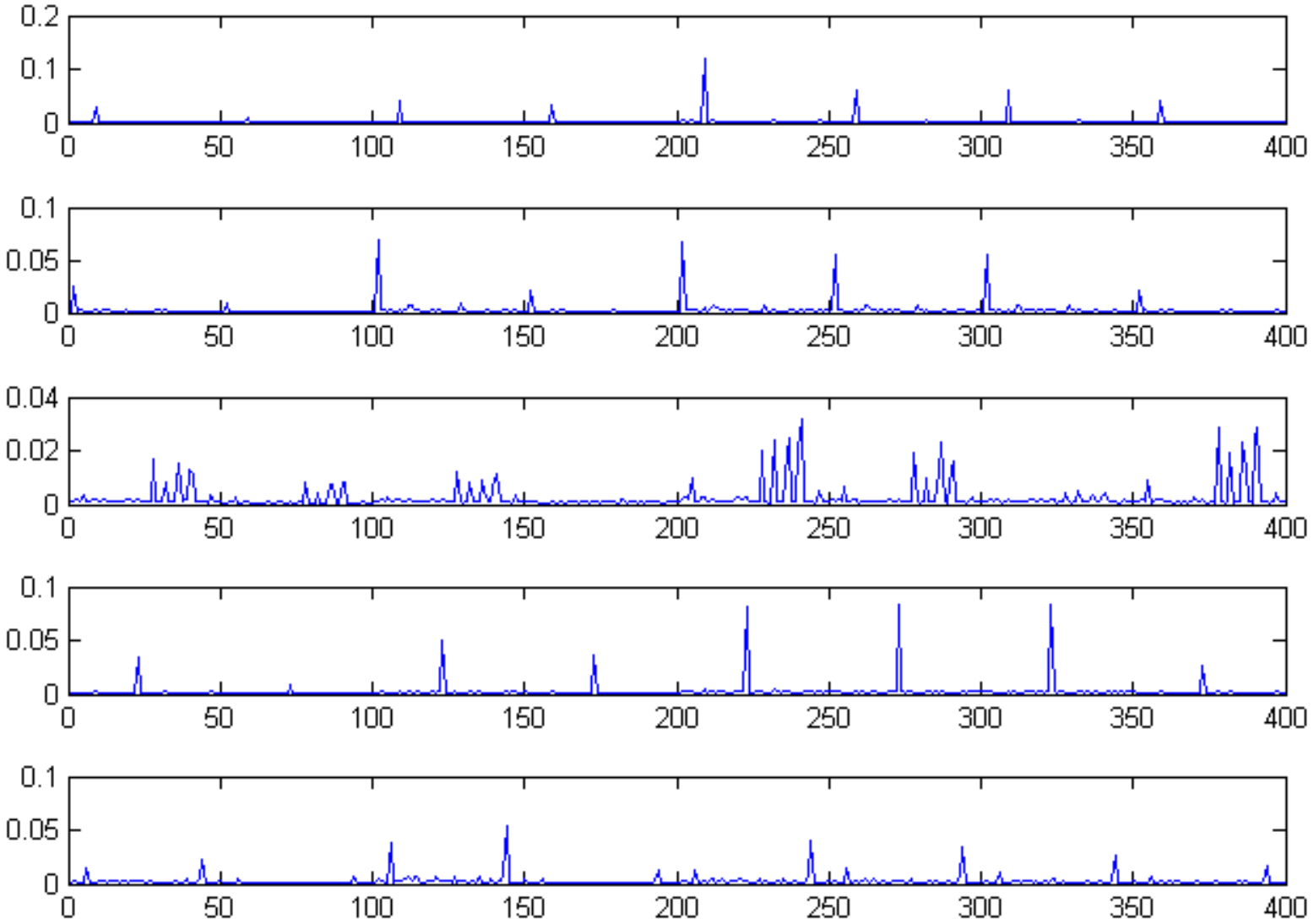
用户类中心点的行为模式分布



（K=5、L=3、C=10）

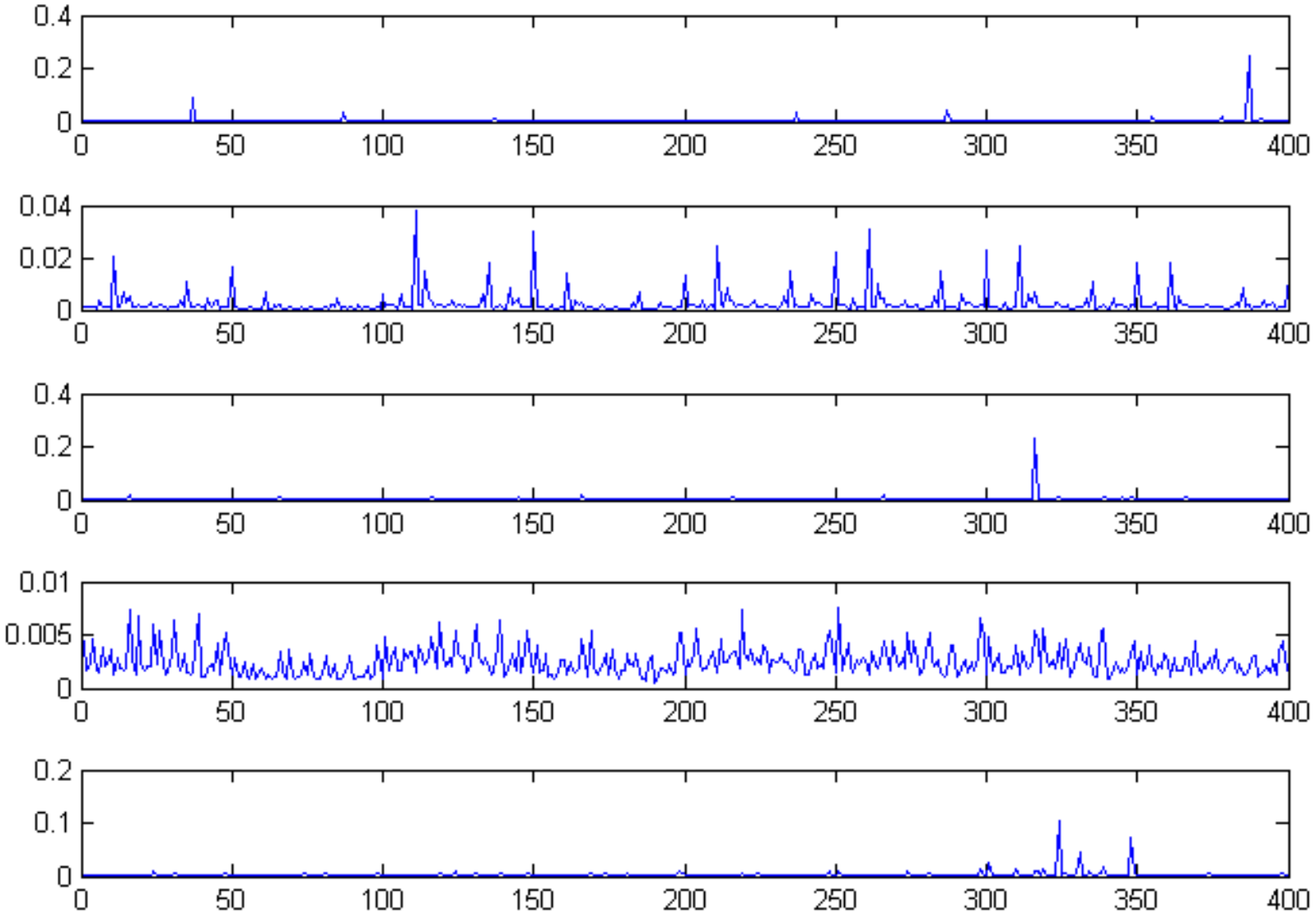
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

用户类中心点的行为模式分布



（K=50、L=8、C=10）

用户类中心点的行为模式分布



（K=50、L=8、C=10）

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Conclusion

- Proposed a coupled LDA model to mine the behavior patterns of IPTV users
- Individual's interests are time-depdent
- The coupled LDA model is also applicable to many other scenarios
  - Seasonality of Taobao purchase behavior
  - Online video services
  - …

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

References

[1] P. Branch, G. Egan, and B. Tonkin. Modeling interactive behavior of a video based multimedia system. In Proceedings of the IEEE ICC, pages 978-982, 1999.

[2] T. Adomkus, R. Bruzgiene, and L. Narbutaite. Influence of users behaviour to iptv service. Electronics and Electrical Engineering, 18(8):105–108, 2012.

[3] T. Qiu, Z. Ge, S. Lee, J. Wang, J. Xu, and Q. Zhao. Modeling user activities in a large iptv system. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, pages 430–441, 2009.

[4] S. H. Hsu, M.-H. Wen, H.-C. Lin, C.-C. Lee, , and C.-H. Lee. Aimed: a personalized tv recommendation system. In Proceedings of the 5th European conference on Interactive TV, 2007.

[5] J. Kim, E. Kwon, Y. Cho, and S. Kang. Recommendation system of iptv tv program using ontology and k-means clustering. Communications in Computer and Information Science, 2011.

[6] R. Konow, W. Tan, L. Loyola, J. Pereira, and N. Baloian. Recommender system for contextual advertising in iptv scenarios. In The 14th International Conference on Computer Supported Cooperative Work in Design (CSCWD2010), pages 617–622, 2010.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, Mar. 2003.

[8] X. Wang and A. McCallum. Topics over time: a non-markov continuous time model of topical trends. KDD’06, p.424–433, 2006.

QUESTIONS?

THANKS!