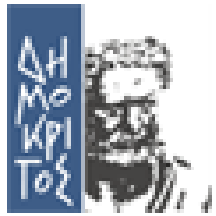
Identifying Personalization Patterns using Intelligent Techniques in XML-based communication records

I. Anagnostopoulos, N. Dimitriou, C. Skianis, and G. Kormentzas

University of the Aegean
Department of Information and Communication Systems Engineering
N.C.S.R. "Demokritos"
Institute of Informatics Telecommunications



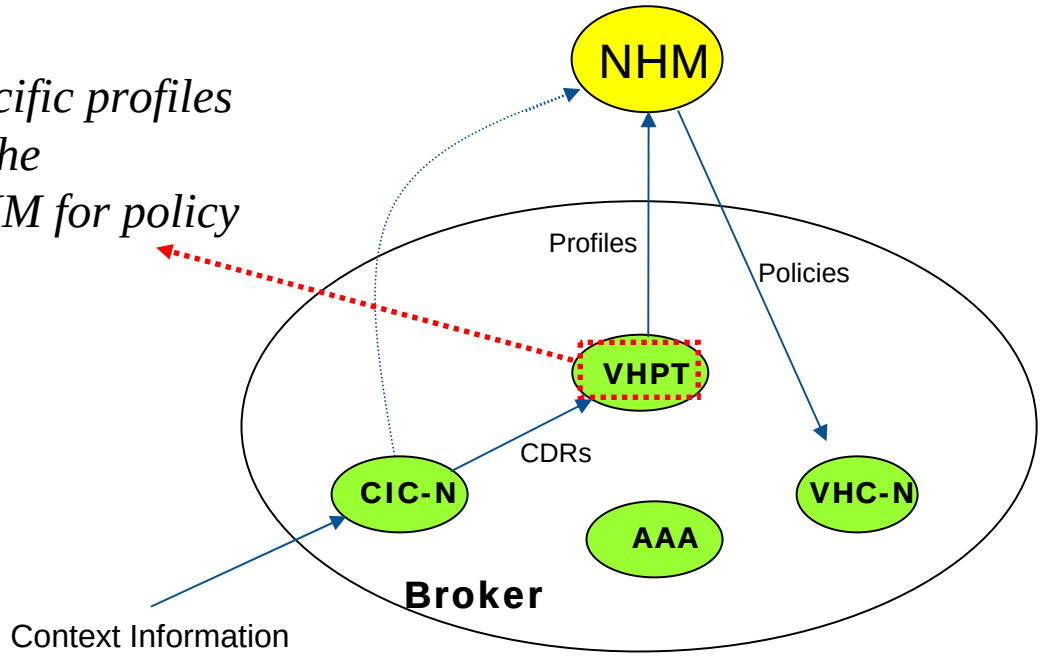
4th Int. Workshop on Semantic Media Adaptation and Personalization, December 14-15, 2009, S. Sebastian, Spain

Presentation overview

This paper deals with…

Data mining and AI methods for identifying or revealing behavioural patterns in large amount of log data in the telecommunication industry (Call Detail Records in Telecommunication Providers)

The VHPT generates the user-specific profiles (regarding the network usage by the subscriber) and sends it to the NHM for policy derivation.



The presented work in this paper is supported by the project INFSO-ICT-216006 HURRICANE (Handovers for Ubiquitous and optimal bRoadband connectIvity among CooperAtive Networking Environments), that is partly funded by the European Union.

Call Detail Records (CDRs)

CDR [*http://en.wikipedia.org/wiki/Call_detail_record*]

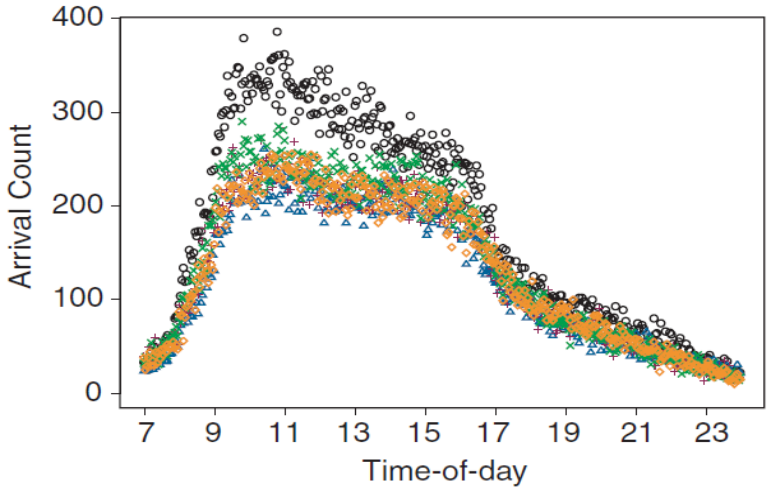
A Call Detail Record (CDR) is the computer record produced by a telephone exchange containing details of a call that passed through it. It is the automated equivalent of the paper toll tickets that were written and timed by operators for long distance calls in a manual telephone exchange.

- the number making the call (A number)
- the number receiving the call (B number)
- when the call started (date and time)
- how long the call was (duration)
- Call Type (0-Voice, A-SMS etc)
- *the number charged for the call*
- *the result of the call (whether it was answered, busy etc)*
- *any facilities used during the call, such as call waiting or call diversion*
- etc…

Data mining in CDRs

e reasons to perform data mining and personalidation provision through CDRs…

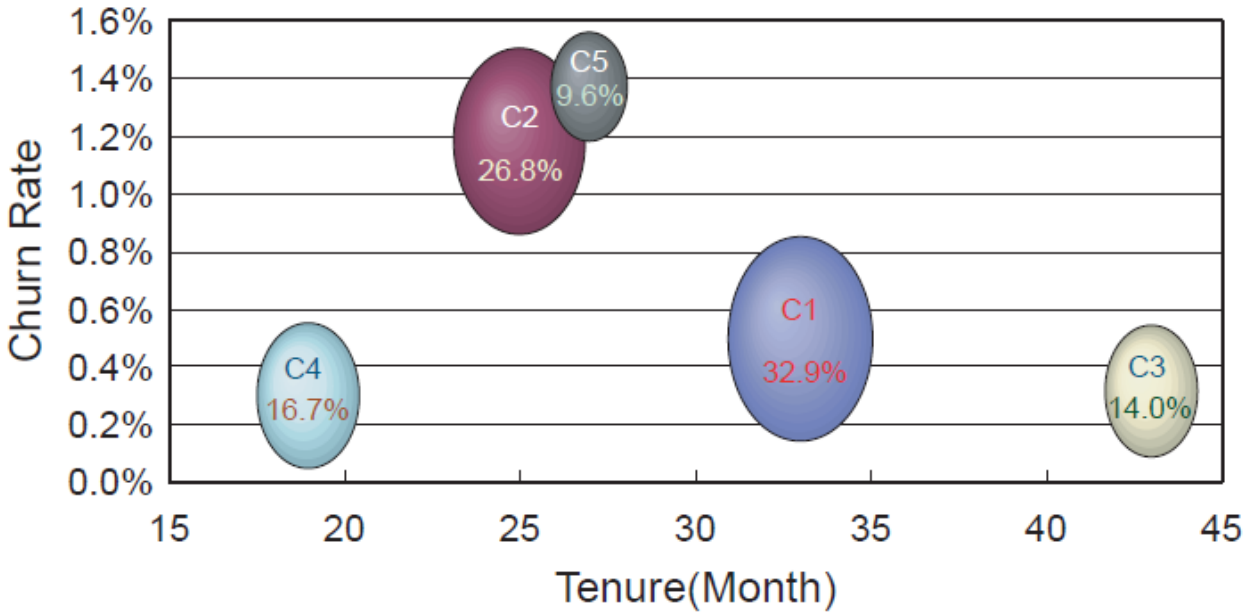
- which customer group is highly profitable , which one is not?
- to which customers should we advertise what kind of special offers ?
- which customers are most likely to churn? (unsubscribe from a provider to another provider)
- how do customer profiles change over time?
- what services differentiate our products from those to our competitors?
-what kind of call rates would increase profit without losing good customers?
- Rare events prediction



VHPT Case Study 1: Churn prediction

Total set: Records from 3583 users in a four-year basis

Data mining lead to five basic clusters along with their centroids



Cluster ID	Tenure	Bill AMT	MOU	MTU	PYMT rate	Percentage of population	Churn rate (%)
C1	H	H	H	H	M	32.9	0.50
C2	L	L	L	L	L	26.8	1.19
C3	H	M	M	M	L	14.0	0.32
C4	L	M	M	M	L	16.7	0.30
C5	M	M	M	M	H	9.6	1.37

VHPT Case Study 1: Churn prediction

Total set: Records from 3583 users in a four-year basis

Data mining lead to five basic clusters along with their centroids

Cluster ID	Percentage of population
C1	32.9
C2	26.8
C3	14.0
C4	16.7
C5	9.6

C1: 1179 users

C2: 960 users

C3: 502 users

C4: 598 users

C5: 344 users

We separated the user populations

We used the mean values of CDRs for 500 randomly chosen users per categories C1 up to C5

Equally trained classes of 500 records each

Total set for training/testing 2500 instances

VHPT Case Study 1: Churn prediction

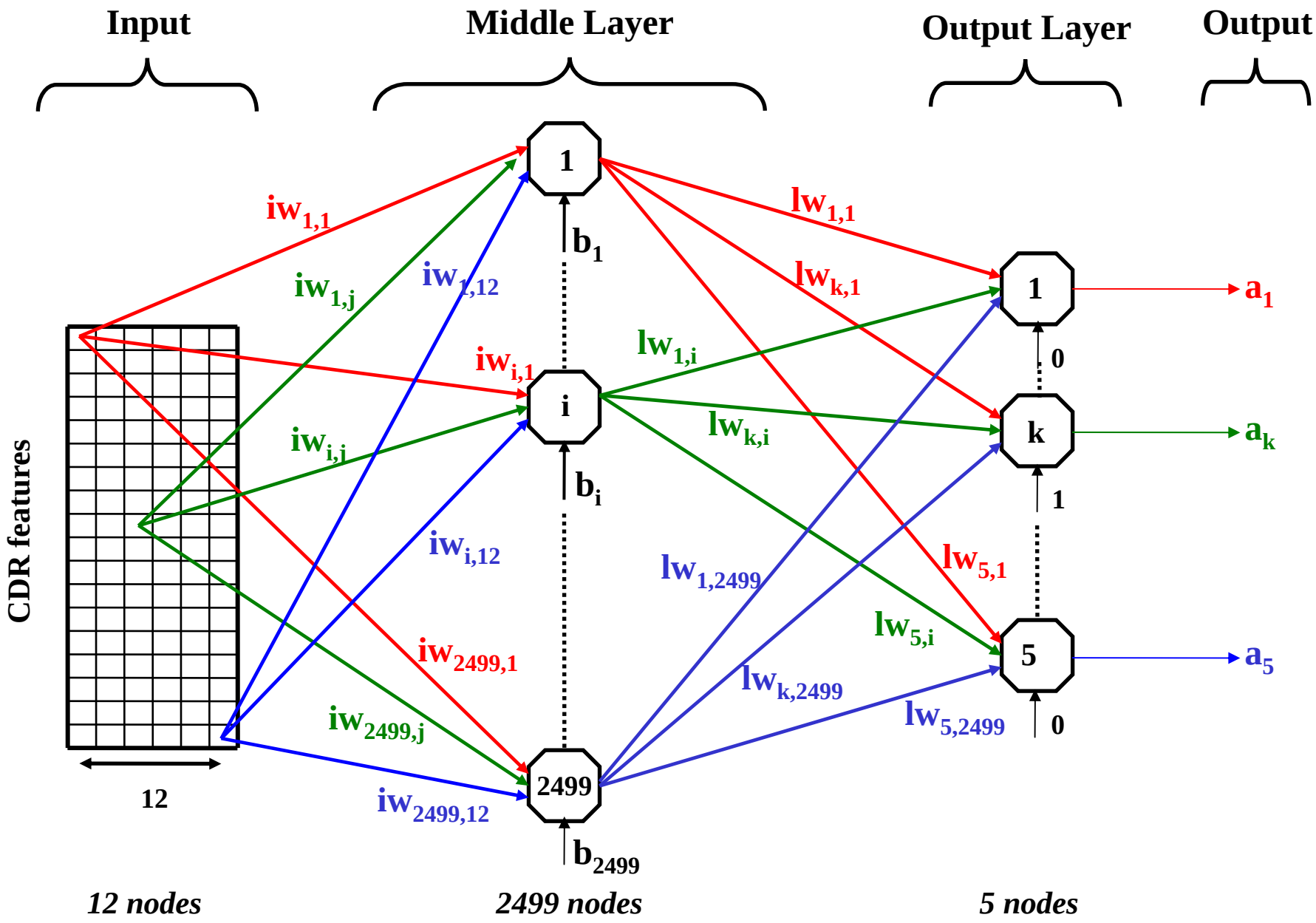
Feature selection: Genetic algorithm

32 → 12 record fields (that best contribute to the categories)

Items	Total customers	Group 1	Group 2	Group 3
ARPU	75.23	379.56	77.96	19.27
MOU	316.65	1122.58	418.25	81.34
Local call percentage	89.52	37.98	**96.04**	**97.58**
Long distance call percentage	4.98	**28.73**	1.67	0.96
IP call percentage	3.33	**9.45**	1.42	0.79
Roam percentage	2.17	**23.84**	0.86	0.67
Idle period local call percentage	10.92	7.89	**17.56**	**15.12**
Idle period long distance call percentage	9.85	6.78	**11.12**	**19.23**
Idle period roam call percentage	7.84	5.71	**10.98**	**18.34**
GPRS traffic volume	0.05	0.15	0.08	0.01
Number of short message	46.09	89.65	**226.34**	6.89
Call diameter	123.73	**287.89**	131.56	38.76

The 12 record fields used (mean values) for the three out of the five classes

VHPT Case Study 1: Churn prediction



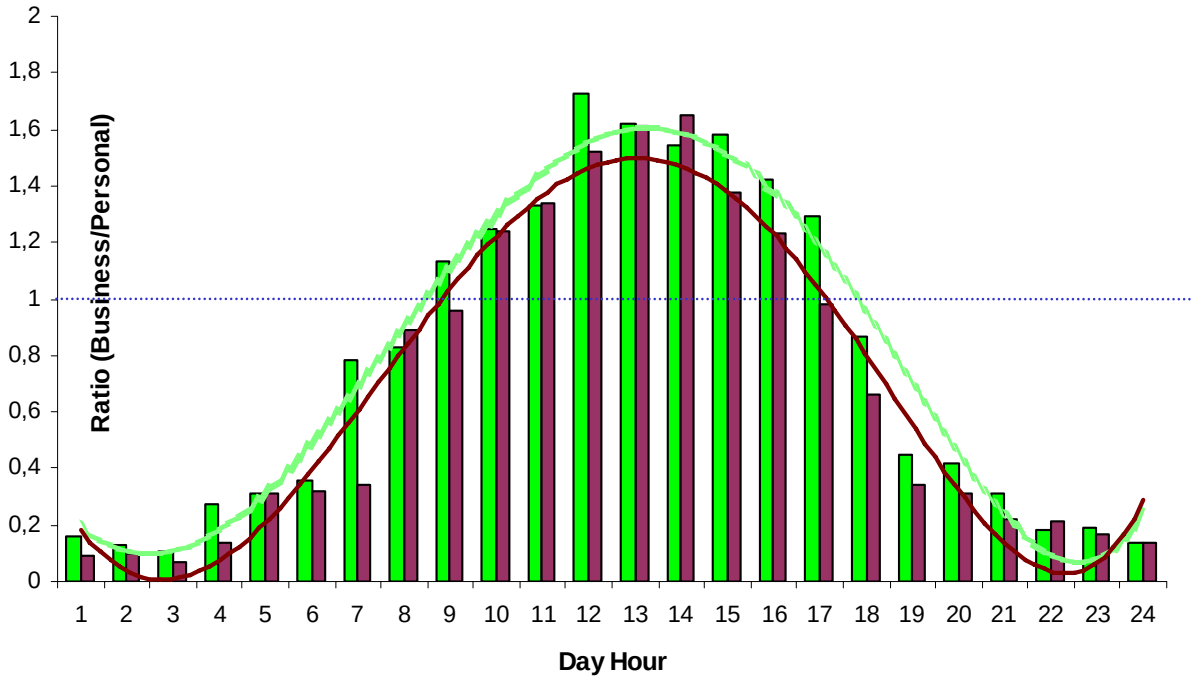
VHPT Case Study 1: Churn prediction

PNN classifier	C1	C2	C3	C4	C5
C1	80,24	5,34	4,15	6,72	3,56
C2	5,27	80,12	4,67	5,07	4,87
C3	3,64	4,66	83,00	4,25	4,45
C4	5,89	6,10	3,46	80,89	3,66
C5	3,92	4,71	5,49	4,12	81,76
Overall Accuracy = 81,20					

ML algorithm	Accuracy
BayesNet	79.30
SVM linear	79.52
SVM nonlinear	74.65
MLP	78.23
DT	77.14
Naive Bayes	65.68
KNN	79.31

Performance of different learning algorithms when the number of clusters is 5

VHPT Case Stuy 2: Monitoring calling behaviour



L: Landlines M:Mobiles		Predicted		
		L	M	Precision (%)
Actual	L	79	12	86,81
	M	10	93	90,29
			Overall performance	88,66

Conclusions / Future work

- The mobile telecommunication marketplace is highly competitive. The marketing managements of mobile telecommunication operators often need to design distinguishable marketing strategy based on different behavior of their mobile customers in order to improve their marketing result.

- Call detail records describe customer utilization behavior. They have more information to describe customer behavior than billing system data.

We presented two case studies based on real CDRs

limitation: hard to found due to privacy issues

Future work consist of real-time personalised services to telecommunication users based on their calling behaviour history

Identifying Personalization Patterns using Intelligent Techniques in XML-based communication records

Thank you for your attention…