Journal of Universal Computer Science, vol. 19, no. 3 (2013), 325-352
submitted: 16/10/12, accepted: 28/1/13, appeared: 1/2/13 © J.UCS

Analysis of Mobile Service Usage Behaviour with Bayesian Belief Networks

Pekka Kekolahti
(Aalto University, Department of Communications and Networking, Espoo, Finland
pekka.kekolahti@aalto.fi)

Juuso Karikoski
(Aalto University, Department of Communications and Networking, Espoo, Finland
juuso.karikoski@aalto.fi)

Abstract: The purpose of this paper is to identify probabilistic relationships of mobile service usage behaviour, and especially to understand the probabilistic relationship between overall service usage diversity and average daily service usage intensity. These are topical themes due to the high number of services available in application stores which may or may not lead to high usage diversity of mobile services. Four analytical methods are used in the study, all are based on Bayesian Networks; 1) Visual analysis of Bayesian Networks to find initially interesting patterns, variables and their relationships, 2) user segmentation analysis, 3) node force analysis and 4) a combination of expert-based service clustering and machine learning for usage diversity vs. intensity analysis. All the analyses were conducted with handset–based data collected from university students and staff. The analysis indicates that services exist, which mediate usage of other services. In other words, usage of these services increases the probability of using also other services. A service called Installer is an example of this kind of a service. In addition, probabilistic relationships can be found within certain service cluster pairs in their usage diversity and intensity values. Based on these relationships, similar mediation type of behaviour can be found for service clusters as for individual services. This is most visible in the relation between System/Utilities and Business/Productivity service clusters. They do not have a direct relationship but usage diversity is a mediator between them. Furthermore, segmentation analysis shows that the user segment called "experimentalists" uses more mediator services than other user segments. Furthermore, "experimentalists" use a much broader set of services daily, than the other segments. This study demonstrates that a Bayesian Network is a straightforward way to express model characteristics on high level. Moreover, Node Force, Direct and Total effect are useful metrics to measure the mediation effects. The clustering implemented as a hybrid of machine learning and expert-based clustering process is also a useful way to calculate relationships between clusters of more than a hundred individual services.

Keywords: handset-based data, Bayesian Networks, machine learning, mobile services, clustering, segmentation
Categories: H.4.0, I.2.6, I.5.1

1 Introduction

Handset-based measurements are a data collection method utilizing smartphones' ability to install third party application software. These measurements are implemented by installing a research application to the mobile phones of opt-in participants. With these measurements rich contextual user level data on mobile service usage can be collected. Handset-based measurements have increasingly been used for a number of purposes in the recent years, applications ranging from sociology [Eagle, 09] to consumer behaviour [Verkasalo, 09]. Mobile service usage in particular has been studied, for instance, by [Verkasalo, 09], [Karikoski, 11] and [Falaki, 10]. A mobile service is a service or application which can be used with a mobile device (in this case a smartphone with Symbian operating system). Mobile service usage was studied from a number of perspectives in the Finnish mobile market with data collected from Symbian devices by [Verkasalo, 09]. Utilizing data collected from 500-700 Finnish smartphone users, [Verkasalo, 09] provided early results of mobile service usage indicating that the handset-based data collection method can be utilized in mobile industry related research and can have real practical relevance. Mobile services were studied in more detail by [Karikoski, 11] by analyzing the interrelationships between mobile communication services with a similar sample as [Verkasalo, 09]. In other mobile platforms, [Falaki, 10] have studied a sample consisting of more than 200 Windows Mobile users. Their main finding based on user interactions, mobile service usage and traffic analysis is that a vast diversity among users exists. The users' behaviour differs with one or more orders of magnitude. Similar results have been obtained also by [Soikkeli, 11]. For a more detailed review of recent handset-based data collection efforts, see [Karikoski, 12].

In our previous research, we have not come across studies, where Bayesian Networks (BN) method has been used for analyzing handset-based data from a holistic mobile service usage perspective. This study uses BN to find service usage patterns and to cluster mobile services, using handset-based measurement data. BN is an analytical method that can be used for inferential analysis, e.g., to make "what if" simulations, to predict behaviour patterns and future trends, to understand why something most probably happened, and to understand which data correlate with other data. It is challenging to analyze the relationships between mobile service usage patterns as the number of available mobile services is very high in the used dataset. It is expected that a BN procedure will offer new easier methods to study this domain. Additionally, handset-based data research area is still new and new insights are expected for using BN. We focus only on methods based on BN and don't qualify the results against other methods like Regression analysis or Neural Networks, because method comparison results already exist, for instance, from [Baesens, 04] and [Verbeke, 12].

Elsewhere in the mobile industry and mobile service related research [Smura, 09] have developed a framework for analyzing mobile service usage, which will be utilized in our research. Mobile services are divided into different categories such as Calling, Messaging, Browsers, Infotainment Clients, Multimedia, Games, Business/Productivity, System/Utilities, Other Applications, and Servers and File Sharing. As the number of mobile services developed for different purposes is exploding with the new mobile platforms and application stores, this framework is valuable for mapping the services together according to some characteristic. Thus, we are able to extract more holistic results from the mobile service usage analysis and utilize the results better, for example, on industry level. Moreover, [Kivi, 09] has identified three dimensions for measuring mobile service usage that will be utilized in our research as well, namely *volume*, *frequency*, and *diversity*. Volume can be measured, for instance, in terms of sessions, duration or bytes. Frequency on the other hand can be measured, for example, as the daily / weekly / monthly intensity of mobile service usage (in the remainder of the article we will use the term *intensity* instead of frequency). Finally, diversity refers to the number of different services used, i.e., how diverse a set of services is being used. Similar constructs used in related literature include *rate of use* and *variety of use* as defined by [Shih, 04]. Rate of use refers to volume and frequency, whereas variety of use refers to diversity.

In this research we will utilize handset-based data collected from 134 Symbian users and analyze it with BN. The main focus is in analyzing the usage of mobile services as a function of the usage of other mobile services. Especially we try to find those services which mediate the usage of other services. Interesting study question is also the possible correlation between two types of service usage patterns, namely diversity and intensity. Moreover, we discuss the generalization, or external validity, of achieved results, as 134 users is still a fairly limited sample size.

2 Methods and data

2.1 Bayesian networks

As Bayesian Networks (also called Bayes Belief Networks (BBN)) is not widely used as an analytical method, an introduction is given to the method, the modelling process and the models. Furthermore, the used metrics are described.

A BN can be created in three ways, namely manually by using expert knowledge, by using machine learning, or a combination of them. As said, a BN is used for inferential analysis (often called predictive analytics), e.g., to predict behaviour patterns and future trends, to make "what if" simulations, to understand why something most probably happened and to understand which data correlate with other data. Some application examples are modelling probabilistic phenomena in risks management domains (e.g., [Lee, 09]), to make driver analysis, classification and clustering tasks for sales and marketing (e.g., [Verbeke, 11]), to elicit and unify expert knowledge for purposes like ecology and business management (e.g., [Kekolahti, 11] and [Kuhnert, 10]), medicine (e.g., [Cruz-Ramirez, 07]), and recently focusing also on telecommunication retention and churn analysis applications (e.g., [Verbeke, 11], [Cinar, 10], [Kisioglu, 11] and [Chang, 10]). BNs started to emerge as an alternative form of probabilistic information representation and probabilistic inference engine through the works of [Pearl, 00] and since then their popularity has increased, first in the academic world and then in business. A good example of increased popularity is a long list of available open source packages and commercial tools, listed by [Anthony, 06]. But still in 2005, regardless of a long list of applications, the main trend in developing analytical models in machine learning was to use regression analysis, neural networks, and decision trees. BNs played only a small role based on comparisons in [Hadden, 07]. The most important features of good methods for inferential analysis are listed by [Verbeke, 12] and [Anderson, 04]. They are predictive accuracy, comprehensibility, justifiability, capability to handle different kinds of data, overall computational performance, and capability to tolerate poor data. BN fulfils those criteria except for computational performance - the complexity increases exponentially as a function of the number of variables and amount of other parameters which slows down the computer performance. Another challenge with BNs comes from the fact, that widely established practices to interpret certain metrics originating from a BN (such as Mutual Information or Node Force) and used in this study do not exist.

A BN is a computational model based on graph and probability theory. The structure of a BN is a Directed Acyclic Graph (DAG). The nodes represent the domain's variables and each arc between nodes represents a probabilistic dependency quantified using a conditional probability distribution table [Barber, 11]. Conditional probability calculation in a BN is based on Bayes' Theorem (Formula 1),

$$P(Y|X) = P(X|Y)P(Y)/P(X), \quad (1)$$

where P(Y|X) is the conditional probability of variable Y given X, also called posterior probability. P(X) and P(Y) are probabilities of X and Y, also called marginal probabilities and P(X|Y) is the conditional probability of X given Y, also called likelihood. A BN topology can be considered as a network of variables connected by the probabilistic dependencies between them and this dependency is maintained by the Conditional Probability Table (CPT) attached to the corresponding variable [van Koten, 06]. According to [Yang, 96], the joint probability distribution, $P(x_1,x_2,\ldots x_n)$ over the BN within variables x_i, where i = 1-n, can be represented as

$$P(x_1,x_2,\ldots x_n) = p(x_i1)p(x_i2|x_i1)\ldots p(x_in|x_i1,x_i2,\ldots x_in\text{-}1). \quad (2)$$

A given joint probability distribution can be represented within different BNs, depending on the preferred node order $(i_1,i_2,\ldots i_n)$. Thus, the BN representation is not unique.

2.1.1 Bayesian Networks learning, clustering and segmentation

The BN learning consists of two phases, namely structural (topology) learning and parameter learning. The structural learning deals with structural relationships between the variables and the output is the BN graph. The parameter learning on the other hand focuses on how the variables quantitatively relate to each other. Three types of algorithms for structural learning exist, unsupervised, supervised, and semi-supervised learning, each of them with their own preferred use cases and desired outcomes. Unsupervised learning is used typically for segmentation and clustering, i.e., when the target is to find hidden structure in the dataset or as an initial analysis phase to study the overall relationships in the dataset. Supervised and semi-supervised learning generate a function or structure that maps input variables to a desired target variable, selected by the expert. Multiple algorithms exist for supervised, semi-supervised, and unsupervised learning [Barber, 11], which have been implemented in multiple software packages [Anthony, 06].

Additionally, the algorithms can be grouped based on model characteristics to causal and optimized structure-methods. Constraints-based algorithms represent causal networks [Pearl, 00] and score-based algorithms represent the optimized structure-grouping [Scutari, 10]. This paper uses score-based methods, where algorithms assign a score to each candidate BN and try to maximize the probability of P(Model|Data) with a selected search algorithm. Different score metrics have been proposed (e.g., [Rissanen, 99], [Yang, 96], and [Yun, 04]). Models in this study are based on the Minimum Description Length (MDL) by [Rissanen, 99]. The MDL principle defines that the best model to represent the data is the smallest which will produce the required accuracy. MDL Score, which is the size of the model minus accuracy of the model, prefers simple BNs over complex ones and is represented as

$$MDL(Model|Data) = |Model|(\log_2 N)/2 - \log_2(P(Data|Model)), \quad (3)$$

where |Model| is the smallest number of parameters to represent all the probabilities and N is the number of data used to calculate the probabilities.

A number of score-based heuristic algorithms for unsupervised learning exist, e.g., Taboo, Taboo Order, Maximum Spanning Tree, EQ (learning of Equivalence classes), SOPLEQ, Hill Climbing, and Grow-Shrink. For supervised learning examples of algorithms are Naïve Bayes, Tree Augmented and Augmented Naïve Bayes, Markov Blanket, and Sons & Spouses ([Barber, 11], [Buntine, 96], [Webb, 05], [Cheng, 01], [Cerquides, 03], [Munteanu, 01], [Jouffe, 01]). The predictive performances of some of the above mentioned algorithms have been tested and also compared against non-BN methods, for example, by [Baesens, 04], [Verbeke, 12], [Cruz-Ramírez, 07] and [Chang, 10]. This study is not focusing on method comparisons but makes a preliminary algorithm comparison within the limits of the existing dataset in order to find the most optimal method for actual studies and for the used dataset.

Cluster analysis is a process, which identifies hierarchical or non-hierarchical groups or clusters of samples that behave similarly or show similar characteristics. Segmentation process on the other hand divides something into pieces (segments) according to selected criteria. This study uses both segmentation and clustering, as part of so called Probabilistic Structural Equation Modelling (PSEM) method, i.e., Bayesian approach to Structural Equation Modelling (SEM), which is discussed, e.g., in [Lee, 07]. In this study we use the term data segmentation when we refer to segmentation of variables' states in order to find an optimum representation of the Joint Probability Distribution defined by all the variables, and variable clustering when we refer to clustering of variables themselves (e.g., [Buntine, 96], [Jain, 99]). The EM (Expectation Maximization) algorithm (e.g., [Pham, 06], [Jain, 99]) is used in this study to calculate the most optimal values for the states of the latent variable (unknown variable) to characterize the manifest variables (a variable that can be observed and measured) both in data segmentation and variable clustering.

2.1.2 Discretization

Discretization of continuous values into numerical intervals is one step in the BN learning process. K-means, Equal frequencies and Equal intervals for unsupervised learning and multiple methods were proposed in [Dougherty, 95], including decision tree for supervised learning. These methods are used in this study when optimizing the learning results by using metrics described in chapter 2.1.3.

The size of the CPT will grow exponentially as a function of the number of parents for a child and the number of intervals of parents and children [Gerssen, 06]. Large CPTs in a BN are the major reason to potentially low computational performance. On the other hand, the number of intervals in variables is an important quality parameter for BN and thus the number is a compromise between quality and computer performance [Uusitalo, 07], [Dougherty, 95]. In addition to quality versus computer performance aspect, a “proper” number of intervals depends on the used machine learning method, number of samples in the dataset, randomness of variables, and also on the purpose of the study [Boulle, 05], [Turner, 06].

In the research presented in this paper the following heuristics are used to select the proper intervals: With maximum spanning tree learning algorithm and three or five intervals for both child and parent, the size of CPT is not higher than nine or 25 cells because the algorithm restricts the number of parents to one. In case of uniform distribution of CPT, 134 samples in the dataset, divided by nine or 25 cells suggest 14.9 or 5.4 observations per cell. Observations of 14.9 with three discretization intervals are above the thumb of rule of minimum 10 samples per indicator, widely used in SEM studies [Westland, 10]. Two and three intervals are selected for this study, because CPT is not in practice uniform and we test also other learning methods, where number of parents might be higher than one. With two parents per child each having two intervals yields 16.8 observations per cell, which we consider reasonable for this study.

2.1.3 Performance metrics for BNs used in this study

Two types of correlation metrics, mutual information and Pearson’s correlation are used for dependency analysis between two individual model variables. Mutual Information I(X,Y) measures how much the observation of random variable Y tells about the uncertainty of X [Barber, 11], [Cover, 01]. In other words it tells how much the information entropy of X is reduced when knowing Y. The entropy of X is written as

$$H(X) = -\sum_{(x\in X)}[p(x)\log_2 p(x)], \tag{4}$$

where x is the possible states of X and p(x) their marginal probability. Mutual information is expressed as

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{5}$$

Mutual information as a nonlinear metric is a benefit but because it is symmetric, we use also Pearson’s correlation to verify the sign of relationship between variables when needed.

The causality theory contains concepts such as direct and indirect effect (see, e.g., [Pearl, 00]). They are described typically with three variables, X, Z, and Y with direct paths between X→Y and Z→Y, and an indirect path between X→Z→Y. Variable Z is called a mediator facilitating the indirect effect between X and Y. Mediation analysis is typically based on the above model of three variables, where causal effects are estimated based on regression or Bayes’ Theorem. These two approaches are compared, for example, by [Chen, 11] and [Zhang, 09]. This study uses the learned BN to estimate indirect and direct effects. The used BN model is not causal but probabilistic, where independence between X and Y given Z can be analyzed from a BN using probabilistic inference. In case of independence, no information is passing between X and Y through Z, formally expressed as $X \perp Y \mid Z$. In a probabilistic BN, a total conditional dependency between X and Y through all possible paths, direct and indirect, is calculated using Formula 2 and the path is said to be not d-separated if dependency exists [Pearl, 00]. On the other hand, d-separated path means that independence between X and Y through Z exists. In that case only direct effects between X and Y, and Z and Y exist. By adapting Pearl's do(x) operator calculus, direct effects for Y can be calculated from the BN. This study uses heuristic metrics for total conditional dependency between X and Y, called Total effect (Te), and for direct relationship between X and Y, called Direct effect (De). Both measure the change of the mean of Y as a function of a small change in the mean of variable X expressed as δ_y/δ_x, where δ_x is a delta of mean of effecting variable X, and δ_y is a delta of mean of target variable Y. In De and Te the relationship is assumed to be linear. Te measures the effect of a variable both as a direct and indirect effect, whereas De holds all the indirect effects fixed (by using adaptation of Pearl's do(x) operator) and thus measures only, as the name says, direct effects. The discarding of intermediating effects in De is based on heuristic Jouffe's Likelihood Matching algorithm, explained by [Conrady, 11b, p.13] as follows: *'As we saw with Pearl's Graph Surgery approach, the core idea of intervention is to set an evidence on the node on which we wish to intervene, while all other ascending nodes remain unchanged. Using this very same idea, we can then intervene on a node by fixing the posterior probability distributions of its covariates*'. De is used in this study to understand causality type of relationships, and Te to understand probabilistic relationships through all the paths, both direct and indirect. These two metrics are especially used to measure mediating (indirect) effects of mobile service usage patterns.

Structural Coefficient (SC) is a practical parameter in structural learning for the control of the MDL score in the used tool. A good initial heuristic value for SC is 1, which prevents typically over-fitting [Conrady, 11a]. The lower the value, the better the accuracy of the learned model but a danger of over-fitting also exists. SC is optimized in this study in a similar way as used in [Conrady, 11a], namely by plotting the Structure/Target's Precision Ratio and selecting the SC-value from elbow-part of the Structure/Target's Precision Ratio -curve.

The dataset contains missing values and our approach is to use EM. EM provides relatively simple, easy-to-implement and efficient tool for this purpose [Enders, 10], [Moon, 96].

The Kullback-Leibler (KL) divergence is used to measure the difference between two marginal probability distributions [Barber, 11], [Cover, 91]. When two marginal probability distributions are equal, D_{KL} is zero. D_{KL} for discrete distributions p = $\{p_i\}$ and q = $\{q_i\}$ of variables x and y is expressed as

$$D_{KL}(p||q) = \sum_i p_i \ln(p_i / q_i). \quad (6)$$

In addition to calculating the differences between marginal probabilities, D_{KL} is used to calculate the differences between joint probability distribution p(x,y) of variables x and y and their marginal probabilities. Intuitively, if x and y are independent, p(x,y) = p(x)p(y) and D_{KL} is zero, no information is gained about y with x. In other words, D_{KL} quantifies the strength of conditional dependency between variables [Schneidman, 03]. D_{KL} is used in this study in three areas: in the variable clustering, data segmentation, and to calculate Node Force. In variable clustering and data segmentation it describes the information gain brought by the manifest variable to the knowledge of latent variable. This means, that it is used to understand the explanatory strength of different manifest variables. Node Force is a heuristic metric provided by the used tool to study the importance of an individual variable within the whole BN. The higher the Node Force per variable is, the more important the variable is in the BN. Node Force (NF) of variable x is calculated as

$$NF_x = \sum_i D_{KL}e_i + \sum_j D_{KL}o_j, \quad (7)$$

where $D_{KL}e$ is KL divergence of entering (towards x) and $D_{KL}o$ KL divergence of outgoing arc of variable x.

2.2 Used BN tool

The selection of BN tools depends on the type of the study, content of the study, and other capabilities. Typical tools used by academia which support also Bayesian methods (e.g., [Weka], [R-project] or [Mahout]) can be seen as machine learning libraries, and they require scripting effort. Hugin, Netica, and Genie (see [Anthony, 06] for more) are mostly manual BN network drawing and analysis tools. We use Bayesialab [Bayesialab], which combines manual drawing of BN, machine learning capabilities and all the reporting possibilities for the purposes of this study.

2.3 Handset-based measurements

With the advent of smartphones it has become possible to install third party application software on mobile phones. Handset-based measurements are a data collection method utilizing this ability. In handset-based measurements a research application is installed in the handsets of the users who have opted in to participate in the measurements. This application collects data on overall usage of the device, including mobile service usage and contextual data. Data on mobile service usage is collected when the service is visible on the foreground of the device's screen. Consequently, processes or services running in the background and not directly interacting with the user are not being analyzed. Moreover, data on mobile communication services, such as, voice calls, SMSs (Short Message Service), MMSs (Multimedia Messaging Service), and email are collected whenever a voice call is made or received, or a message is sent or received. The handset-based data collection process and data utilized in this research are documented in detail in [Karikoski, 12].

2.3.1 Data characteristics

The original dataset was collected from 200 users in the OtaSizzle project of Aalto University during 2009 and 2010 [Karikoski, 12]. In this study we have only utilized data collected from Symbian devices from users that have produced data for a minimum of 30 days. Thus, we are left with 134 users with males and females represented by 91 and 16 users, respectively (37 users were unidentified). The average participant is born in 1985 (median 1988). The participants in the measurements have been identified in [Karikoski, 12] as early adopters of mobile devices and services, and are biased towards Finnish male students in their early twenties. Close to 600 unique services used by the participants have been thoroughly analyzed and mapped to categories according to the framework of [Smura, 09]. In this process, the non-user generated, mostly operating system related service events have been excluded from the data, in order to ensure that the service usage events depict user behaviour.

As in most computer generated data collection efforts, the data are displayed on user-level in a time stamped log form. In this study, the data have been aggregated from single service usage events and pre-processed by calculating the daily service usage intensities for each user by dividing the number of service usage events of a given user by the number of days that that user has produced data in the measurements. With services such as voice calls, the volume of usage as the length of voice calls has been measured. The average length of the incoming / outgoing voice calls is then aggregated on user level. As additional variable the diversity of mobile service usage has been calculated for each user as the number of services used. In case no usage of a given service at all per person exists, the value has been put to zero.

The average service usage intensity follows a long tail form in Figure 1, where most used services are traditional voice call and SMS services, contacts, clock, calendar, music, email, maps, photos, and internet browsing. It could be analyzed if a model such as Zipf's law applies by using, for instance, the tools described in [Naldi, 03]. However, in the scope of this chapter we will only present the long tail to describe the characteristics of the data and not perform further analysis. From the most used services messaging and message editor are specific for the Symbian platform and need to be elaborated. These services are both used in conjunction with text messages and email, and it is not possible to identify which one of the services was used together with the messaging or message editor service.

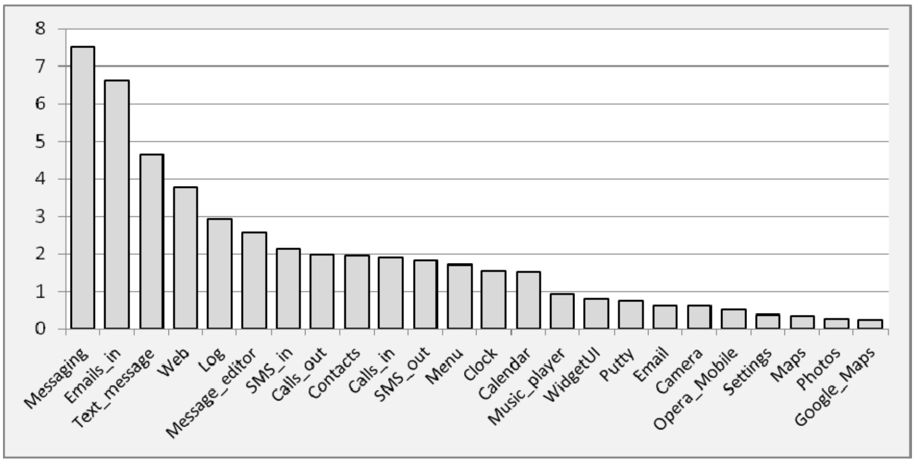

Figure 1: Average service usage intensity as usage times per day of most popular 24 services out of 571 services.

Figure 2 describes mutual information values between two variables (between a parent and a child) when the model has been created with unsupervised learning and by using maximum spanning tree algorithm and SC of 0.4. SC was selected based on the elbow-part of the Structure/Target's Precision Ratio -curve. The created model (Figure 3) includes all the 571 daily service usage intensities, and diversity, birth year, and gender as variables. With the learning process and by using SC of 0.4, the relationships between some of the variables are so weak, that no arcs between them and other variables exist. In that case, they can be interpreted to be independent from the other variables. From 571 service usage intensity related variables 506 are dependent (thus connected variables). Figure 3 is the actual model, which indicates visually where the parent-child connections are with the highest mutual information. The variables inside dotted circles represent services used only by one or some persons. Mutual information between these kinds of variables is small, typically 0.05-0.07, which can be interpreted nearly as independent behaviour. Independent behaviour means that the daily usage of a certain variable doesn't correlate with the daily usage of another variable. Visual inspection indicates that diversity obviously plays an important role in the whole model. It is easy also to recognize where the 22 strongest mutual information values of Figure 2 are located in the model. The rest of variable pairs in the model (blue circles) have smaller mutual information. The BN graph layout algorithm creates a symmetric graph and also utilizes the arc length and the repulsive power between the variables to enhance the visual effect and clarity of the BN graph.

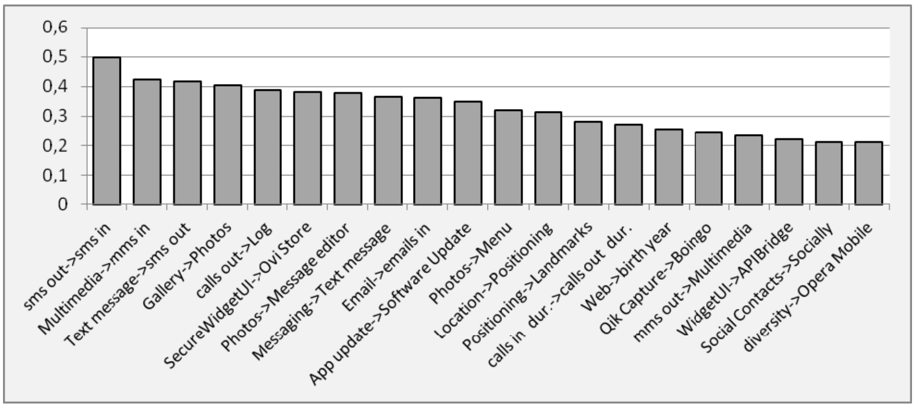

Figure 2: 20 strongest mutual information values between parent-child variables in a BN model of 506 variables.

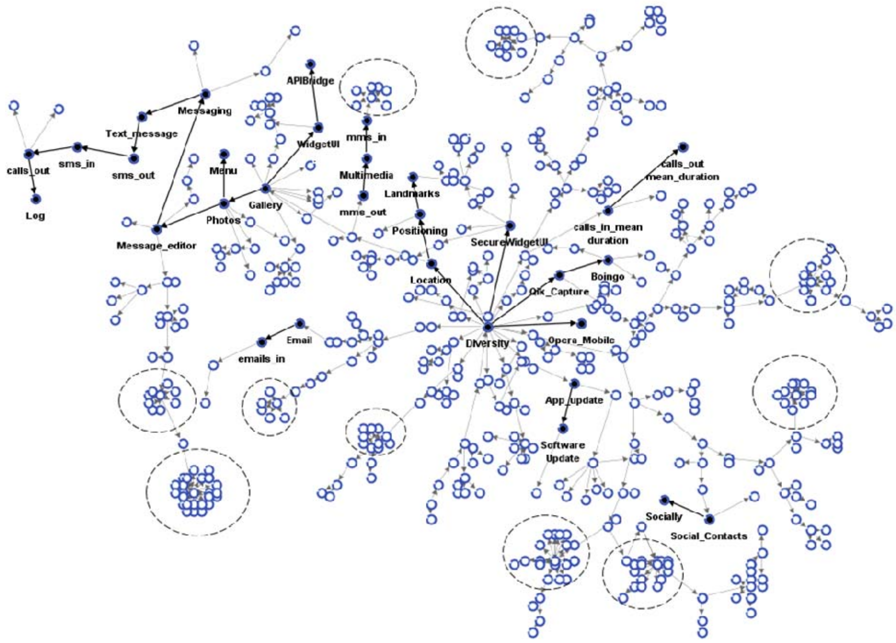

Figure 3: BN created with unsupervised learning using maximum spanning tree algorithm. The small circles denote the variables, the arc between two variables denotes an existing dependency, measured, e.g., as mutual information.

3 Results

3.1 Importance of individual services in service portfolio

The usage intensity reflects the importance of individual services to the end users. The higher the average usage intensity of a service is, the more important the service is in the daily life of a user. In addition to this kind of importance, a service can also enable or affect the usage of other services. Node Force metrics were used to analyze local structures in the BN model, and to verify, whether clear structures can be found which describe enabler roles of services. Taboo learning and SC 0.42 were used to find all the possible relationships for each variable. Maximum spanning tree algorithm used in the model in Figure 3 fits well for visual inspection of a BN, because it simplifies the BN structure by restricting the number of parents per a child to one.

Figure 4 describes 18 most important services from Node Force point of view. Also the local connection to other nearest services is visible. These connections in fact contribute to the Node Force metric. The figure shows that some of the services, e.g., Photos, Installer, Settings, App update, SecureWidgetUI, and Boingo have low intensity but at the same time the Node Force value belongs to the 18 highest. They all are connected to five or more other services. Thus, their importance measured as Node Force within the whole BN model is high.

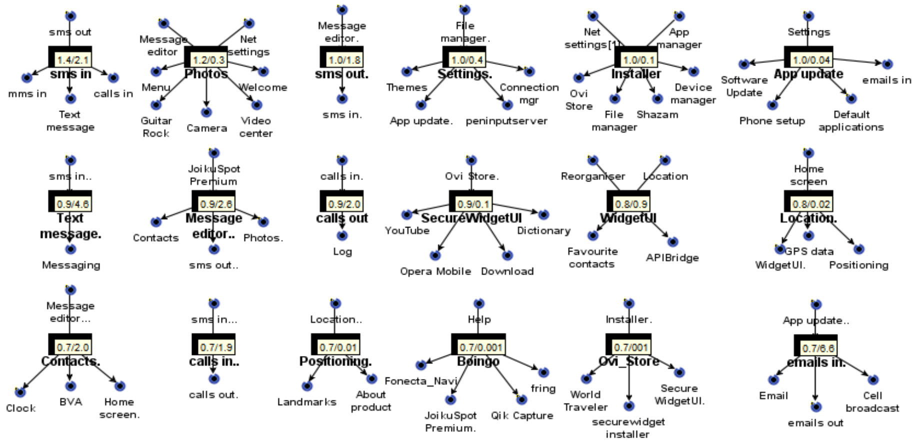

Figure 4: Node Force versus intensity of 18 services denoted as A/B, where A = Node Force and B = intensity. The services are locally connected with 2-7 other services.

Table 1 lists the 18 services of Figure 4 as a function of intensity, Node Force, and effect of individual variable intensity to the whole model's mean intensity, with and without indirect effects. Indirect effects have been avoided by fixing all other model's variables except the target; see the computational model in Appendix. Figure 4 and Table 1 indicate that some services are clearly enablers for others as they affect the overall intensity mean of the model even if they have not been used themselves intensively. We call these kinds of services mediators as they mediate usage of other services. Important indicators of mediators include the number of links to other services in BN network, and a high mutual information value between services. The clearest examples of variables with a high number of links are Ovi Store, Installer, Photos, SecureWidgetUI, and Boingo and of the high mutual value (where number of links is small) are SMS out, Text messages, and Calls in. SMS in represents both categories, because it has both many links to other services and high mutual information with those services.

3.2 Segmentation of users based on their gender, age and diversity

As can be seen from the long tail in Figure 1, the majority of services were used only seldom. The usage intensity value was zero, when the service is not used by persons at all during the study period. The dataset did not have missing values, but the entropy between variables with a low amount of usage is small even if Pearson's correlation coefficient might be high. In segmentation analysis the zero intensity values were interpreted as missing values in order to make sure that the segmentation is based on the actual usage and not non-usage. Missing values above 25% per service were not included in the learning process. In total 24 services out of 571 fulfilled these criteria.

The first step in the user segmentation process was to use data segmentation to segment the service usage intensities to an optimum number and level to represent (summarize) the intensities of each 24 services. In the BN the latent variable (the variable in the middle of the model in Figure 5) represents the data segmentation results. The number of states, i.e., the number of intensity segments was defined as the average states of all the manifest variables, i.e., three. Call duration, diversity, age, and gender were discarded from the first step.

Service name	**Node Force**	**Service Intensity**	**Model's Delta mean, Conditional Intensity, Fixed**	**Model's Delta mean, Conditional Intensity, Non fixed**	**Diff Non Fixed-Fixed**
SMS in	1.3603	2.143	0.009	0.064	0.055
Photos	1.136	0.302	0.002	0.01	0.008
Installer	1.0329	0.124	0.003	0.006	0.003
SMS out	1.0026	1.826	0.01	0.063	0.053
App update	0.9983	0.043	0	0.05	0.05
Settings	0.9713	0.371	0.01	0.011	0.001
SecureWidgetUI	0.9433	0.095	0.001	0.006	0.005
Message editor	0.94	2.595	0.019	0.06	0.041
Calls out	0.9334	1.977	0.01	0.041	0.031
Text message	0.8541	4.634	0.017	0.062	0.045
WidgetUI	0.813	0.861	0.008	0.011	0.003
Location	0.7684	0.015	0	0.003	0.003
Positioning	0.7217	0.004	0	0.002	0.002
Calls in	0.718	1.899	0.005	0.034	0.029
Contacts	0.7119	1.948	0.007	0.027	0.02
Boingo	0.702	0.001	0.003	0.01	0.007
Ovi Store	0.7014	0.014	0	0.07	0.07
Emails in	0.6873	6.618	0.075	0.079	0.004

Table 1: Node Force, Service intensity, conditional delta mean of the model given max minus min intensity of target service with indirect effects being fixed, conditional delta mean of the model given max minus min intensity of target service, and delta of the two values.

Figure 5 represents the BN after the data segmentation process. The latent variable in the middle of the BN has three states representing the intensity characteristics of the user segments S1, S2, and S3. Service usage intensity mean values for those segments as well as their probabilities are seen in the Figure 5.

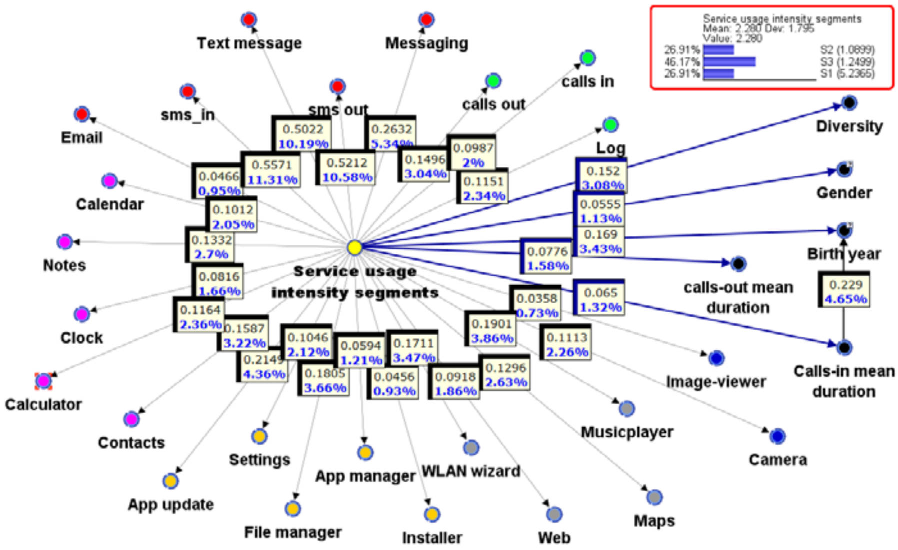

Figure 5: BN learned with data segmentation. "Service usage intensity segments" is the latent variable representing the 24 manifest variables. The black numbers are Kullback-Leibler divergence values (D_{KL}) between the manifest and latent variables, and blue numbers represent the relative significance of the connection.

Figure 6 depicts the characteristics of user segments S1, S2, and S3. The segments are called "heavy communicators", "experimentalists", and "basic smartphone users", respectively. Characteristics of each segment are the following:

- **S1**: "a heavy communicator" has a high daily SMS usage intensity (3.97 for in and 3.69 for out), highest intensity for Calls in and out (2.42 and 2.79, respectively) and also for Contacts (2.88). Moreover, a somewhat higher usage intensity of Clock (1.92) and Notes (0.17) exists.
- **S2**: "an experimentalist" has the highest daily intensity for Music player (1.65), Camera (0.77), Email (0.87), WLAN wizard (0.24), Notes (0.26), Maps (0.56), App manager (0.08), and Installer (0.18), which fit to the usage profile of a person who wants to experiment and use the phone for many purposes. Furthermore, usage of SMS is moderate.
- **S3**: "a basic smartphone user" has tried the phone with many features but uses it moderately, mostly for Calls (in 1.68, out 1.55), but also sometimes to look at the Emails (0.43) and Images (0.45).

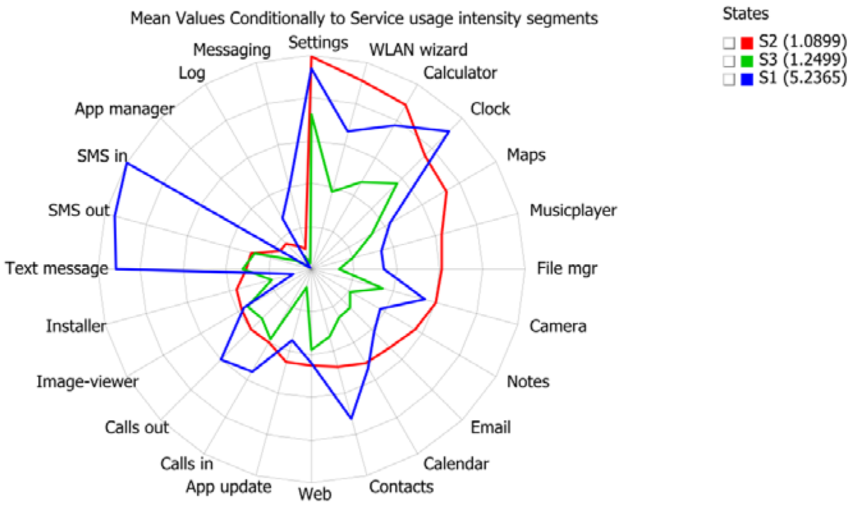

Figure 6: Service usage characteristics of user segments S1, S2, and S3 analyzed as normalized conditional intensity means. Normalized conditional intensity is zero in the middle of the chart and maximum 100% on the rim.

The second step in the segmentation process was to segment the users according to their age, gender, service usage diversity, and call duration to S1, S2, and S3. Taboo learning was used, and in addition to the variables birth year, diversity, gender, duration of incoming calls, and duration of outgoing calls, only the variable service usage intensity was used, and all its manifest variables in Figure 5 were omitted.

Figure 7 highlights the segmentation results from the second step as probabilities of each segment given posterior data for diversity, gender, birth year, and duration of incoming and outgoing calls. Birth years were classified to 1954-1976 (10.19% of users), 1977-1986 (33.77% of users) and 1987-1991 (59.04% of users). Conclusions are:

- Those persons who have low diversity (who have used all in all less than 21 services out of nearly 600) belong with 80.71% probability to segment S3 (“basic smartphone user”). Persons who have used 71 or more services belong with 66.26% probability to segment S2 (“experimentalists”).
- Women belong with 53.57% probability to segment S1 (“heavy communicators”) and only with 6.81% probability to segment S2 (“experimentalists”), whereas men’s segment distribution is not so sharp.
- Users in 1954-1976 category belong with 97.85% probability to segment S3, whereas users in the 1987-1991 category belong rather evenly to all three segments. Low averages in incoming and outgoing call durations are linked with 53.53% and 97.62% probabilities to segment S3, but also medium incoming call duration with 95.84% probability is linked to segment S3.

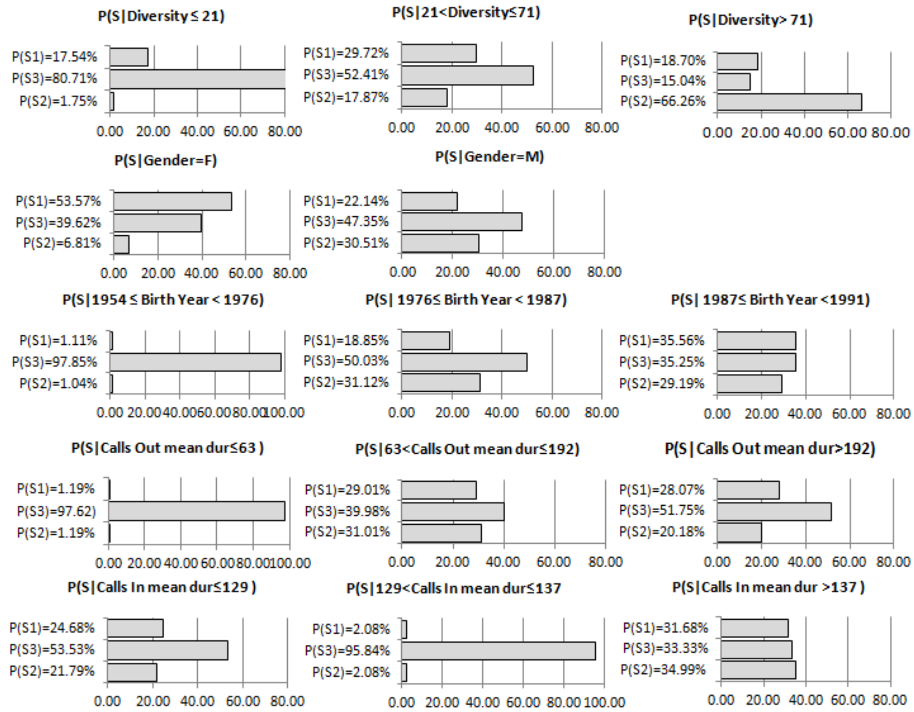

Figure 7: Probabilities of segments S1 ("heavy communicators"), S2 ("experimentalists"), and S3 ("basic smartphone users") given posterior data for diversity, gender, birth year, duration of incoming calls (in seconds), and duration of outgoing calls (in seconds).

3.3 Expert-based clustering

Expert-based clustering in this study means using expert's knowledge in clustering the services, and exploiting the generated clusters to calculate conditional probabilities between the clusters acquired with unsupervised learning process with the Taboo method. The expert-based clustering of services is based on the framework of [Smura, 09] and was performed by Karikoski. The individual services were first identified by means of their Symbian UIDs, and then the services were clustered according to their common characteristics. The services were divided into 10 clusters: Business/Productivity (75 services), Games (114 services), System/Utilities (106 services), Messaging (32 services), Calling (9 services), Multimedia (51 services), Infotainment Clients (51 services), Servers and File Sharing (5 services), Browsers (7 services), and Other Applications (121 services). According to our information, a hybrid clustering of mobile services based on the combination of expert knowledge and machine learning has not been applied before.

Figure 8 is the cluster network after Taboo learning. SC was set to 0.15 in order to find all possible, even noisy relationships between variables. The model contains diversity, birth year and gender in addition to the 10 clusters, in order to find their relationships with the clusters. The name of the latent variable in the model (like Games or Multimedia) describes the characteristics of each manifest variable included in a certain cluster and its states describe quantitatively (summarize) the characteristics of the same manifest variables. It is visible from the BN in Figure 5, that diversity plays a central role in the model because it is connected to all other variables in the BN. In another words, diversity correlates with intensity represented by clusters. Clusters in the model have 2-4 states depending on the number of states of the manifest variables. Moreover, 12 cluster to cluster relations are visible in Figure 8, presented as dotted lines. Node Force analysis indicates that indeed diversity has the strongest node force (2.29), then Multimedia (1.18), Other Applications (1.05), System/Utilities (0.67), and Messaging (0.60). All the services listed in Table 1 are included in these four mentioned clusters.

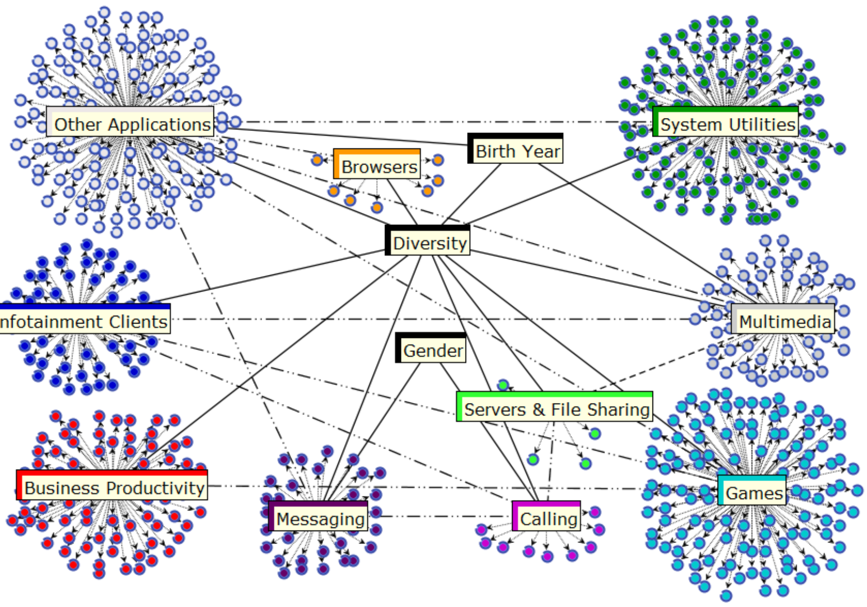

Figure 8: A BN consisting of ten expert-based service intensity clusters and diversity, gender, and birth year. The labels of individual services per each cluster have been omitted due to clarity.

Table 2 summarizes the relationships between each latent variable and gender, diversity, and birth year. The relationship is measured as mutual information and as Pearson's correlation. The table also contains p-values for each parent-child pairs, which indicate, that in 16 out of 28 parent-child connections the results are not statistically significant (within 5% p-value). This was expected due to the fact of low intensities in the majority of services and low number of samples. Thus, conclusions from the clustering are only indicative, representing the used dataset. Figure 8 and Table 2 demonstrate that both System/Utilities and Business/Productivity have the strongest relationship with diversity measured as mutual information. They are not directly related based on Figure 8, but indirectly through diversity. Thus, diversity acts as an intermediate variable between them. Diversity has also similar but weaker relationships with Multimedia, Other Applications, and Infotainment Clients clusters. Furthermore, Multimedia has direct relationships with the two others.

Parent→Child	**Mutual Information**	**Pearson's Correlation**	**p-value**
Diversity→System/Utilities	0.6037	0.5715	0
Multimedia→Other Applications	0.5208	-0.7832	0
Diversity→Business/Productivity	0.4278	0.5209	0
Diversity→Multimedia	0.3083	0.3971	0
Diversity→Infotainment Clients	0.3738	0.5075	0
Infotainment Clients→Games	0.134	0.391	0
Multimedia→Infotainment Clients	0.1551	0.4309	0
Diversity→Other Applications	0.1527	-0.3546	0.00001
Other Applications→System/Utilities	0.1218	-0.3828	0.000012
Multimedia→birth_year	0.1311	0.2578	0.000089
Diversity→Messaging	0.1953	0.2506	0.000144
Diversity→Games	0.1037	0.2575	0.000474
Other Applications→birth_year	0.0861	-0.3218	0.00057
Other Applications→Messaging	0.0926	-0.2741	0.000595
Games→Business/Productivity	0.0425	0.2482	0.004211
Servers & File Sharing→Multimedia	0.0348	0.2111	0.036161
Servers & File Sharing→Calling	0.047	0.3095	0.057471
Infotainment Clients→Calling	0.0163	0.146	0.079786
Messaging→Calling	0.0365	0.1735	0.093033
Messaging→gender	0.0601	-0.2597	0.109426
Diversity→birth_year	0.082	0.0908	0.124613
Diversity→Browsers	0.0189	0.1492	0.351465
Browsers→Other Applications	0.0046	0.0792	0.364406
Other Applications→Games	0.0034	0.068	0.378412
Calling→gender	0.0003	-0.0196	0.68102
Diversity→Servers & File Sharing	0.0099	-0.1086	0.688547
Diversity→Calling	0.0031	0.0268	0.788317

Table 2: Mutual information, Pearson's correlation, and p-value of each connection in Figure 8.

Even if general conclusions cannot be drawn, Te was analyzed from the point of view of acquiring indicative results (Figure 9). In Te analysis, the three non-clustered variables, birth year, gender, and diversity, and the Messaging, Multimedia, System/Utilities, Infotainment Clients, Calling, and Business/Productivity clusters were selected as target variables, each in turn. Te was calculated between the target and other variables. Descriptive terms "slightly", "moderately", "strongly", and "very strongly" are used for Te in the following way: $0 \leq$ slight Te< 0.25, $0.25 \leq$ moderate Te < 0.5, $0.5 \leq$ strong Te< 0.75, and $0.75 \leq$ very strong Te ≤ 1.00.

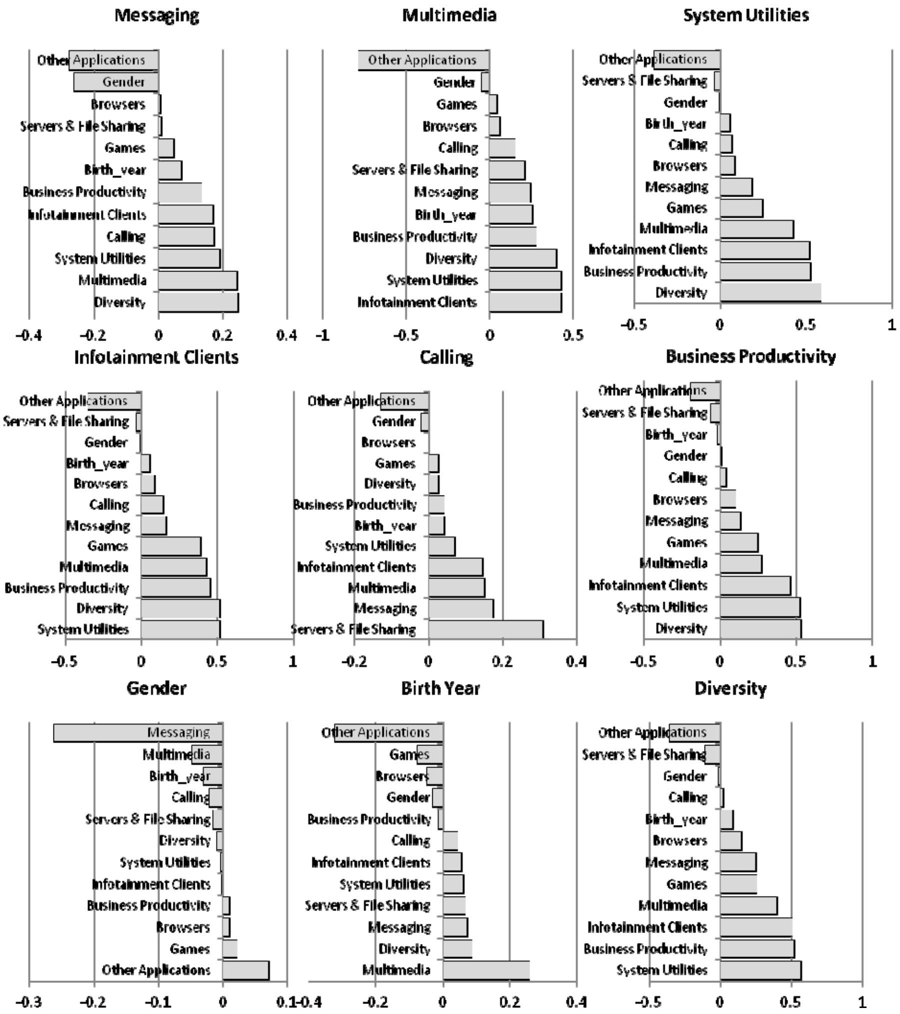

Figure 9: Te has been calculated from the cluster network including the three non-clustered variables gender, birth year, and diversity.

A few remarks from Figure 9 are:

- A slight Te with the Messaging cluster and diversity, Multimedia, and System/Utilities exists.
- A moderate Te with Multimedia cluster and Infotainment Clients, System/Utilities, and diversity exists.
- A strong Te is found with the System/Utilities cluster and diversity, Business/Productivity, and Infotainment Clients.

- Infotainment Clients cluster has a strong Te with System/Utilities and diversity, and a moderate one with Business/Productivity, Multimedia, and Games.
- Calling cluster has a moderate Te with Servers and File Sharing, and a small one with Messaging, Infotainment Clients, and Multimedia.
- Business/Productivity has a strong Te with diversity, System/Utilities, and Infotainment Clients, and a moderate one with Multimedia.
- Diversity has a Te value among the top three within all the cases where the service cluster has been set as a target except for the Calling cluster. Gender has a strong Te with the Messaging cluster, but otherwise slight. Birth Year has a strong Te with Multimedia cluster.

In order to understand in more detail the relationships between diversity and clusters, as well as diversity, birth year and gender, a comparison between Te and De was done in Figure 10. In each of the relationships the absolute value of Te is clearly higher than that of De. Some of the Te and De values are negative, meaning that when diversity has increased, intensity has decreased. Same is visible also in Figure 9. As Figure 10 indicates, System/Utilities, Business/Productivity, and Infotainment Clients have the largest positive Te with diversity as a target.

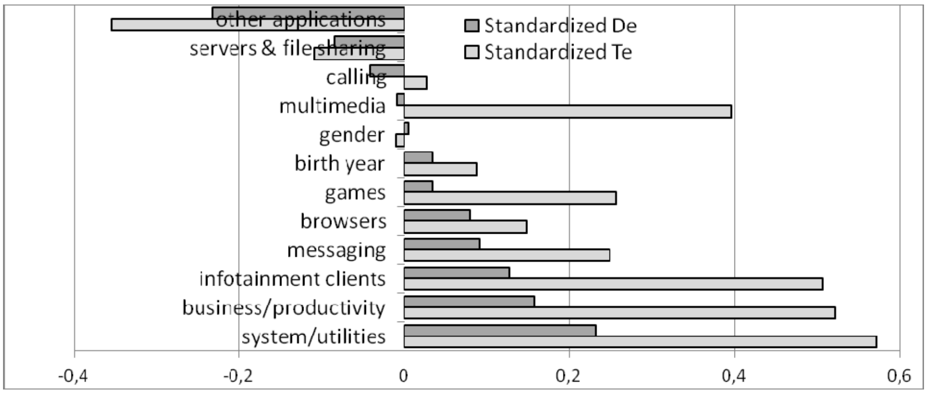

Figure 10: Te and De between diversity (the target) and intensity of other variables.

4 Discussion

The results of the study can be divided into four parts:

- BN as a visual representation communicates efficiently the relationships between variables.
- Some mobile services act as mediators relaying the usage intensity of other services.

- The user segments and their relationships with birth year, gender, and diversity demonstrate the differences in daily usage of services within each segment.
- A relationship between diversity and intensity of usage is identified by using expert-based service clustering and machine learning. Moreover, relationships between service clusters and gender, birth year, and diversity are identified.

The tools, methods, models, and metrics used in this article are all related to Bayesian procedures. Motivation to this was to exploit some of the Bayesian method's capabilities, e.g., a possibility to look at the models in a graph format, to utilize some of the less used (and partially heuristic) metrics like mutual information, Node Force, Direct effect and Total effect. One motivation was also to benefit from Bayesian method's possibility to tolerate poor data.

Figure 3 is an example of how a BN can visually highlight the existing statistical relationships and also their strengths even when the number of variables is very high, in hundreds. For example, clumps of variables are visible in Figure 3, which tell that intensity of those services is low and they are most probably used only by one or a few users. Similarly there seems to be a strong statistical relationship between some variables, e.g., Menu & Photos, Positioning & Location, Log & Calls out, and Calls in duration & Calls out duration. And lastly, diversity most probably acts as an important variable in the whole model due to the high number of direct relationships with many other variables in the BN. A BN created with unsupervised learning offers a quick way to pre-study the dataset and to find potentially important or key variables and statistical relationships also in a complex environment thus saving research time.

In the in-depth analysis services were grouped qualitatively to two groups – mediators and non-mediators. The grouping was based on the services' intensity and Node Force values. Based on Figure 4 and Table 1, Ovi Store and App update are strong mediators, because regardless of their low intensity, the mediator role is visible clearly from relatively high delta non fixed minus fixed conditional mean intensity values. Calls in and Emails in have both high intensity but they are not mediators. SMS in has a high intensity value and at the same time it is a clear mediator. Emails in is used daily frequently, obviously users are checking the incoming emails. However, interestingly this activity does not mediate Emails out at all. One potential reason for this might be the lack of usability of the Email service. Node Force analysis according to our information has not been used before to rank services to mediators and non-mediators. This kind of ranking and metrics offers a new tool and insights for service planners and handset manufacturers. They can rank the services to non-mediators, mediators, and both, and managers can pay special attention to the activities which could increase the intensity of mediator services (e.g., in terms of usability), in order to increase the daily usage of other services.

The 134 users consisting of 91 males and 16 females were segmented based on their service usage intensity to three segments, "heavy communicators", "experimentalists", and "basic smartphone users". The descriptive names for segments were given afterwards to describe the service profiles in each segment. The "experimentalists" use a broad set of services daily, including the mediator services, as seen from Figure 6 and Table 1. In fact, nearly 75% of users use their Symbian devices as a normal feature phone and do not benefit of its other capabilities remarkably. Two potential reasons for this might be the low importance of the service for the user, and low usage intensity of mediator services, e.g., due to lack of their usability. The segmentation process itself yielded logical results within Symbian devices, and can be used similarly for any platform to find typical service segments per platform.

A conditional probability relationship between diversity and intensity is an interesting question in the mobile service ecosystem. In other words, what kind of a statistical relationship does there exist between a highly or a lowly diverse a set of services being used, versus intensity measured as average daily usage times per service. Nowadays application stores have hundreds of thousands of services available. This kind of a setting might easily increase the diversity of service usage supposing that applications enabling this are easy to use (and thus good mediators). But does this kind of setting boost average daily usage and how?

The framework of [Smura, 09] offered a way to categorize hundreds of services into practical clusters for machine learning based modelling. To our knowledge, this is the first time, when the procedure, where expert-based clustering and machine learning is used to analyze impact of diversity behaviour to intensity behaviour and vice versa. The role of diversity based on the results is central within many clusters. It is especially so with System/Utilities, Business/Productivity, Multimedia, and Infotainment Clients. The results then mean, that user behaviour where many services and applications are used is reflected also as high daily service usage, but not with all services. Based on differences between Te and De in Figure 10, in the end the diversity-intensity relationship seems to be indirect more than direct. A clear example is the Multimedia cluster - if diversity increases, it has a positive but indirect effect to daily usage of multimedia services.

The indicative conclusion is that when the number of services in the application stores increases, obviously services will be downloaded more which increases diversity. Within used dataset, the high diversity behaviour means also high daily usage of services especially belonging to clusters of System/Utilities, Business/Productivity, Multimedia, and Infotainment Clients. The phenomenon is especially visible with System/Utilities and Business/Productivity according to Table 2 and Figure 9. No direct relationship between Business/Productivity and System/Utilities exists based on the cluster BN. However, an indirect relationship exists through diversity and other clusters. Thus, diversity creates an indirect link between System/Utilities and Business/Productivity clusters. As services included in System/Utilities are used before the services included in Business/Productivity, assumption is that the former acts as a mediator cluster for the latter. System/Utilities contains all the capabilities that the Symbian operating system offers to users to configure the phone. Consequently, one could hypothesize that if the daily usage of these services is high, then the diversity of usage and the daily usage of Business/Productivity services are also high. Moreover, if a user has low diversity, obviously he/she does not configure, change, personalize, or update his/her phone as much as a high diversity user. Even if this study has not analyzed it separately, all the 75% of users belonging to segment S1 and S3 potentially are such kind of users.

This study faced computational challenges due to the high number of variables in the BN, and the high number of parents per a child in Taboo learning. The third reason to computational challenges is typically too high a number of states per a variable, but they were optimized in this study to two and three. Also other types of optimization were used, such as avoiding too low numbers for the SC parameter, and restricting the number of parents to one (in maximum spanning tree learning) in certain analyses. A danger exists, that these kinds of actions yield high bias with low variance, and thus special care is needed in the optimization. Luckily more efficient computers with more available memory as well as new and optimized algorithms are introduced to reduce the challenge.

Analysing mobile service usage behaviour with the sample used in this article yields many potential threats to the external validity of the results. As discussed, the sample of 134 mobile phone users is still fairly limited, and biased towards innovators and early adopters of mobile phones and services. When collecting data from advanced handsets, such as the ones used in this research, the main challenge remains that the users tend to be more technology-savvy than the general population. Moreover, the observations presented in this article are restricted to the Symbian platform. The expert-based clustering of mobile services also inevitably brings some subjectivity to the observations, but categorization frameworks such as the one by [Smura, 09] help in doing this consistently. Thus, we have not presented the usage analysis results of this article as representative of the general population, but rather as observations representing the used dataset.

A natural future research area would be to update the dataset with current situation with Google Android, Apple iOS and Windows Phone devices. Furthermore, regardless of the target audience, the number of samples should be in order of magnitude higher to increase the credibility and external validity of the results. Samples, which would enable comparing different communities, would be interesting as well. Another relevant topic would be to study how much the diversity of services will increase if the number of applications in the application store is small, e.g., 10000 compared to a situation where it is, e.g., 500000. Similarly, the question can be extended to cover the need of increase in mobile network capacity as a function of diversity. This study analyzed non-mediating and mediating services, but not why some services are mediators and why others are not. This would be one research topic for the future as well.

5 Conclusions

We have studied mobile service usage measured as daily usage intensity and diversity using Bayes Network based modelling with information theory based metrics like mutual information, and heuristic metrics such as Node Force, Te, and De. In addition, a unique hybrid clustering procedure was applied, where expert-based mobile service cluster definitions have been used in an unsupervised learning process to find dependencies between clusters of services (measured as their daily usages), as well as between the clusters and the overall usage diversity of services.

The study categorized services to two groups, mediators and non-mediators, based on their intensity and Node Force values, as well as their overall effect to the average intensity of all services. Intensity and Node Force analysis provided the list of mediator services, whereas their ranking order was determined by calculating the target service's effect to the average intensity of all services. Ovi Store, Installer, Photos, SecureWidgetUI, and Boingo are examples of mediator services, and Emails in is an example of a non-mediator service.

User segmentation based on daily service usage produced three segments, "heavy communicators", "basic smartphone users", and "experimentalists". The two first use the phone mostly for calls and SMSs whereas an "experimentalist" uses a broader service set daily including mediator services. The users with birth year 1976 or earlier will mostly belong to the "basic smartphone user" segment. Nearly 54% of females belong to "heavy communicators" and less than 7% to "experimentalists".

Service clustering analysis demonstrates a strong relationship between diversity of usage and the clusters of System/Utilities, Business/Productivity, Infotainment Clients, Multimedia, and Other Applications. System/Utilities cluster is assumed to be a mediator cluster for Business/Productivity due to common strong mutual information with diversity. However, a direct relationship is lacking.

This study demonstrates with an example, that BN is an easy to understand way to express model characteristics on a high level. The Node Force as well as Direct effect and Total effect proved to be useful metrics within the used dataset to measure mediation effects between variables. Moreover, segmentation brings additional information about service usage behaviour. Finally, expert-based clustering data combined with unsupervised learning created a new BN, based on which relationships between intensity and diversity can be measured.

Even if we treat the results as qualitative, they bring service planners and handset manufacturers some indicative information, which can be used in the service planning process to enhance service's success. Furthermore, the new methods and metrics can be used to define the roles of services in the service portfolio.

Acknowledgement

The work has been supported by the OtaSizzle and MoMIE research projects and the Future Internet Graduate School FIGS. The authors wish to thank MobiTrack Innovations Ltd. for providing the mobile audience measurement platform, and Matti Aksela and Heikki Hämmäinen for providing comments to the manuscript. The sponsoring from Nokia and Elisa to this work is also acknowledged.

References

[Anderson, 04] Anderson, R., Mackoy, R., Thompson V., Harrell G.: "A Bayesian Network Estimation of the Service-Profit Chain for Transport Service Satisfaction"; Decision Sciences, 35(4), 2004, 665-689.

[Anthony, 06] Anthony, K.: Introduction to Causal Modeling, Bayesian Theory and Major Bayesian Modeling Tools for the Intelligence Analyst, USAF National Air and Space Intelligence Center, 2006, http://www.au.af.mil/au/awc/awcgate/nasic/intro-causal-modeling.pdf

[Baesens, 04] Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., Vanthienen, J.: Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers, European Journal of Operational Research, 156(2), July 2004, 508-523.

[Barber, 11] Barber, D.: Bayesian Reasoning and Machine Learning, Cambridge University Press, 2011, http://www.cs.ucl.ac.uk/staff/d.barber/brml

[Bayesialab] Bayesialab software, http://www.bayesia.com/

[Boulle, 05] Boulle, M.: Optimal Bin Number for Equal Frequency Discretizations in Supervized Learning, Intelligent Data Analysis, 9(2), March 2005, 175-188.

[Buntine, 96] Buntine, W.: A guide to the literature on learning probabilistic networks from data, IEEE Transactions on Knowledge and Data Engineering, 8(2), April 1996, 195-210.

[Cerquides, 03] Cerquides, J., Mantaras, R.: Tractable Bayesian Learning of Tree Augmented Naive Bayes Models, In Proceedings of the 20th International Conference on Machine Learning, 2003, 75-82.

[Chang, 10] Chang, K., Yong, N.: Bayesian Network Approach to Predict Mobile Churn Motivations: Emphasis on General Bayesian Network, Markov Blanket, and What-If Simulation, Future Generation Information Technology, LNCS 6485, 2010, 304-313.

[Cheng, 01] Cheng, J., Greiner, R.: Learning Bayesian Belief Network Classifiers: Algorithms and System, Lecture Notes in Computer Science, 2056, 2001, 141-151.

[Chen, 11] Chen, C.: Bayesian Analyses of Mediational Models for Survival Outcome, Doctoral Dissertation, University of Cincinnati, Cincinnati OH, USA, July 2011.

[Cinar, 10] Cinar, D., Kayakutlu, G.: Scenario analysis using Bayesian networks: A case study in energy sector, Knowledge-Based Systems, 23(3), April 2010, 267-276.

[Conrady, 11a] Conrady, S., Jouffe, L.: Breast Cancer Diagnostics with Bayesian Networks, March 2011, http://www.conradyscience.com/white_papers/WBCD_V7.pdf

[Conrady, 11b] Conrady, S., Jouffe, L.: Causal Inference and Direct Effects, September 2011, http://www.conradyscience.com/images/white_papers/causal_inference_v16.pdf

[Cover, 91] Cover, T.M., Thomas, J.: Elements of information theory. Wiley, New York, 1991.

[Cover, 01] Cover, T.M., Thomas, J.: Entropy, relative entropy and mutual information, Elements of Information Theory, John Wiley & Sons, October 2001.

[Cruz-Ramírez, 07] Cruz-Ramírez, N., Acosta-Mesa, H.G., Carrillo-Calvet, H., Nava-Fernandez, L.A., Barrientos-Martinez, R.E.: Diagnosis of breast cancer using Bayesian networks: A case study, Computers in Biology and Medicine, 37(11), November 2007, 1553-1564.

[Dougherty, 95] Dougherty, J., Kohavi, R., Sahami, M.: Supervised and Unsupervised Discretization of Continuous Features, In Proceedings of the 12th International Conference on Machine Learning, July 1995, 194-202.

[Eagle, 09] Eagle, N., Pentland, A., Lazer, D.: Inferring Social Network Structure using Mobile Phone Data, Proceedings of the National Academy of Sciences (PNAS), 106(36), July 2009, 15274-15278.

[Enders, 10] Enders, C.K.: Applied Missing Data Analysis. The Guilford Press, April 2010.

[Falaki, 10] Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R., Estrin, D.: Diversity in Smartphone Usage, In Proceedings of the 8th International Conference on Mobile systems, applications, and services, June 2010, 179-194.

[Gerssen, 06] Gerssen, S.G., Rothkrantz, L.J.M.: Bayesian Network Learning for Rare Events, In Proceedings of the International Conference on Computer Systems and Technologies, June 2006.

[Hadden, 07] Hadden, J., Tiwaria, A., Roy, R., Ruta D.: Computer assisted customer churn management: State-of-the-art and future trends, Computers and Operations Research, 34(10), October 2007, 2902-2917.

[Jain, 99] Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review, ACM Computing Surveys, 31(3), September 1999, 264-323.

[Jouffe, 01] Jouffe, L.: New Search Strategies for Learning Bayesian Networks, In Proceedings of the 10th Int. Symposium on Applied Stochastic Models and Data Analysis, June 2001.

[Karikoski, 11] Karikoski, J., Luukkainen, S.: Substitution in smartphone communication services, In Proceedings of the 15th International Conference on Intelligence in Next Generation Networks, October 2011, 313-318.

[Karikoski, 12] Karikoski, J.: Handset-based data collection process and participant attitudes, International Journal of Handheld Computing Research, 3(4), October-December 2012, 1-21.

[Kekolahti, 11] Kekolahti, P.: Using Bayesian Belief Networks for Modeling of Communication Service Provider Businesses, In Proceedings of the 8th Bayesian Modelling Application Workshop, July 2011.

[Kisioglu, 11] Kisioglu, P., Topcu, I.: Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey, Expert Systems with Applications, 38(6), June 2011, 7151-7157.

[Kivi, 09] Kivi, A.: Measuring mobile service usage: methods and measurement points, International Journal of Mobile Communications, 7(4), 2009, 415-435.

[Kuhnert, 10] Kuhnert, P., Martin, T., Griffiths, S.: A guide to eliciting and using expert knowledge in Bayesian ecological models, Ecology Letters, 13(7), May 2010, 900-914.

[Lee, 07] Lee, S.Y.: Structural equation modelling: A Bayesian approach. John Wiley & Sons Ltd., January 2007.

[Lee, 09] Lee, E., Park, Y., Shin, J.: Large engineering project risk management using a Bayesian belief network, Expert Systems with Applications, 36(3), April 2009, 5880-5887.

[Mahout] Mahout library, http://mahout.apache.org/

[Moon, 96] Moon, T.K.: The expectation-maximization algorithm. IEEE Signal Processing Magazine, 13(6), November 1996, 47-60.

[Munteanu, 01] Munteanu, P.: The EQ Framework for Learning Equivalence Classes of Bayesian Networks, In Proceedings of the IEEE International Conference on Data Mining, 2001, 417-424.

[Naldi, 03] Naldi, M.: Concentration indices and Zipf's law, Economics Letters, 78(3), March 2003, 329-334.

[Pearl, 00] Pearl, J.: Causality: Models, Reasoning, and Inference, Cambridge University Press, 2000.

[Pham, 06] Pham, D.T., Ruz, G.A.: Unsupervised training of Bayesian networks for data clustering. Proceedings of the Royal Society A, 465(2109), July 2009, 2927-2948.

[R-project] R software, http://www.r-project.org/

[Rissanen, 99] Rissanen, J.: Hypothesis Selection and Testing by the MDL Principle, Computer Journal, 42(4), 1999, 260-269.

[Schneidman, 03] Schneidman, E., Still, S., Berry, M.J., Bialek, W.: Network information and connected correlations, Physical Review Letters, 91(23), December 2003.

[Scutari, 10] Scutari, M.: Learning Bayesian Networks with the bnlearn R Package, Journal of Statistical Software, 35(3), 2010, 1-22.

[Shih, 04] Shih, C.F., Venkatesh, A.: Beyond Adoption: Development and Application of a Use-Diffusion Model, Journal of Marketing, 68, January 2004, 59-72.

[Smura, 09] Smura, T., Kivi, A., Töyli, J.: A framework for analysing the usage of mobile services. Info, 11(4), 2009, 53-67.

[Soikkeli, 11] Soikkeli, T., Karikoski, J., Hämmäinen, H.: Diversity and end user context in smartphone usage sessions, In Proceedings of the 5th International Conference on Next Generation Mobile Applications, Services and Technologies, September 2011, 7-12.

[Turner, 06] Turner, C.W., Lewis, J.R., Nielsen, J.: Determining Usability Test Sample Size, International Encyclopedia of Ergonomics and Human Factors, 2006, CRC Press, 3084-3088.

[Uusitalo, 07] Uusitalo, L.: Advantages and challenges of Bayesian networks in environmental modelling. Ecological modelling 203(3-4), May 2007, 312-318.

[van Koten, 06] van Koten, C., Gray, A.R.: An application of Bayesian network for predicting object oriented software maintainability, Information and Software Technology, 48(1), January 2006, 59-67.

[Verbeke, 11] Verbeke, W., Martens, D., Mues, C., Baesens, B.: Building comprehensible customer churn prediction models with advanced rule induction techniques, Expert Systems with Applications, 38(3), 2011, 2354-2364.

[Verbeke, 12] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B.: New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, European Journal of Operational Research, 218(1), April 2012, 211-229.

[Verkasalo, 09] Verkasalo, H.: Handset-Based Analysis of Mobile Service Usage, Doctoral Dissertation, Helsinki University of Technology, Espoo, Finland, April 2009.

[Webb, 05] Webb, G., Boughton, J., Wang, Z.: Not So Naive Bayes: Aggregating One-Dependence Estimators, Machine Learning, 58(1), January 2005, 5-24.

[Weka] Weka software, http://www.cs.waikato.ac.nz/ml/weka/

[Westland, 10] Westland, J.C.: Lower bounds on sample size in structural equation modelling, Electronic Commerce Research and Applications, 9(6), November-December 2010, 476-487.

[Yang, 96] Yang, S., Chang, K.C.: Comparison of Score Metrics for Bayesian Network learning, In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, October 1996, 2155-2160.

[Yun, 04] Yun, Z., Keong, K.: Improved MDL Score for Learning of Bayesian Networks, In Proceedings of the 2nd International Conference on Artificial Intelligence in Science and Technology, November 2004, 98-103.

[Zhang, 09] Zhang, J., Wedel, M., Pieters, R.: Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis, Journal of Marketing Research, 46(5), October 2009, 669–681.

Appendix - calculation of columns 4 and 5 in Table 1

Definitions for network and variables:

G is a BN consisting of n variables N_i with k states s_j; $v_j = v(N_i = s_j)$ is the mean of state s_j

Nt is the target variable, s_{min} is Nt's minimum, and s_{max} maximum interval (state)

Ns_j are the m surrounding variables of Nt (connected to Nt)

Definition of the probabilities:

$P(N_i)$ is the posterior probability of N_i

$p_j = p(N_i=s_j)$ is the probability of the state s_j and $P(N_i)=[p_1,p_2,p_3,...p_k]$

$P_{init}(Ns_j)$ is the initial probability of Ns_i without evidence on the network

$P_u(N_i)$ is the posterior probability of N_i without evidence on the network

$P_f(N_i)$ is the posterior probability of N_i with the probability distribution of Ns_j fixed to $P_{init}(Ns_j)$ and can be expressed as $P_f(N_i) = P(N_i|P(Ns_1) = P_{init}(Ns_1),...,P(Ns_m) = P_{init}(Ns_m))$

Definition of the means:

$M(N_i)$ is the mean of N_i

$M(N_i)= \sum_{j=1}^{k} p_j * v_j$ and

$M(G) = (\sum_{i=1}^{n} M(N_i)) / n$ is the mean of the model G

Calculation of column 4 "model's conditional delta mean intensity, fixed" in Table 1:

For each N_i $P_{fmin}(N_i) = P(N_i|Nt = s_{min}, P(Ns_1) = P_{init}(Ns_1),...,P(Ns_m) = P_{init}(Ns_m))$ and

$P_{fmax}(N_i) = P(N_i|Nt = s_{max}, P(Ns_1) = P_{init}(Ns_1),..., P(Ns_m) = P_{init}(Ns_m))$

$M_{fmax}(N_i)$ can be computed from $P_{fmax}(N_i)$ and $P_{fmin}(N_i)$ from $P_{fmin}(N_i)$

We define $M_{fmax}(G)$ to be the mean of the model when $Nt = s_{max}$ and $P(Ns_j) = P_{init}(Ns_j)$,

and $M_{fmin}(G)$ is the mean of the model when $NT=s_{min}$ and $P(Ns_j)=P_{init}(Ns_j)$

Thus $\Delta_f = M_{fmax}(G) - M_{fmin}(G)$

Calculation of column 5 "model's conditional delta mean intensity, non fixed" in Table 1:

For each N_i we can write $P_{umin}(N_i) = P(N_i|Nt = s_{min})$ and $P_{umax}(N_i) = P(N_i|Nt = s_{max})$

$M_{umax}(N_i)$ can be computed from P_{umax} and M_{umin} from P_{umax}

We define $M_{umax}(G)$ as mean of the model when $Nt = s_{max}$ and $M_{umin}(G)$ when $Nt = s_{min}$

Thus $\Delta_u = M_{umax}(G) - M_{umin}(G)$