content

Hybrid recommender System - Context aware

Data:

Usage (BB/VOICE/PEO TV)

* Location
* Usage - package wise

\*\* User demographics

* + (AGE group/ Gender/)
  + Bill payment methods (card/banks/cash)
  + Payment pattern (debt or delayed payments/ Pay on time)
  + Sentiment Scores on
    - \*\*CRM logs - inquiries/ complains - text
    - \*\*Review extracted from Social Media (on relevant services/packages)

Multi class classification -> ensemble model to choose best classifier depend on data

**Approach of Binning -> rules followed**

* Bins should be all the same size. For example, groups of ten or a hundred.
* Bins should include *all* of the data, even [outliers](https://www.statisticshowto.com/statistics-basics/find-outliers/). If your outliers fall way outside of your other data, consider lumping them in with your first or last bin. This creates a rough [histogram](https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/histogram-make-chart/) —make sure you note where outliers are being included.
* Boundaries for bins should land at whole numbers whenever possible (this makes the chart easier to read).
* Within the range (max-min)

**My approach: Related to spread of data (skewness)**

**Putting into bins considering inter quartile ranges (25%, 50%, 75%, and the highest usage groups)**

## Freedman-Diaconis’s Rule

This formula uses the [interquartile range (IQR)](https://www.statisticshowto.com/probability-and-statistics/interquartile-range/):

**2(IQR)n−1/3**

[https://stats.stackexchange.com/questions/143438/optimal-number-of-bins-in-histogram-by-the-freedman-diaconis-rule-difference-be\](https://stats.stackexchange.com/questions/143438/optimal-number-of-bins-in-histogram-by-the-freedman-diaconis-rule-difference-be%5C)

**Data collection methodology:**

The data dump or raw data will be collected from the

CDR’s for the telecommunication company, Structured

Query Language (i.e. SQL) procedures will be used to

collect the SQL dump or raw CDR’s will be processed using

big data techniques, this data will be split into training set,

test set and validation set.

**Comparing Recommender models over their accuracy or performance:**

**Confusion Matrix**

A confusion matrix illustrates the accuracy of the solution to a classification problem.Given n classes a confusion matrix is a m x n matrix, where Ci,j indicates the number of tuples from D that were assign to class Ci,j but where the correct class is Ci . Obviously, the best solution will have only zero values outside the diagonal a confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. Some standards and terms:

1. True positive (TP): If the outcome from a prediction is p and the actual value is also p, then it is called a true positive.

2. False positive (FP): However, if the actual value is n then it is said to be a false positive.

3. Precision and recall: Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

**Modeling recommenders: statistical approach:**

\*\* Coefficients:

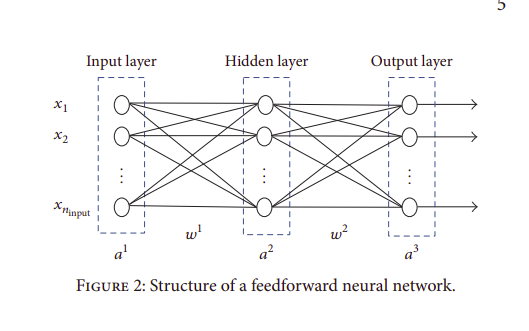
Pearson

Spearman  
\*\* Correlations:

Cosine

Cosine Similarity

**\*\* ANN - how it works**

****

Artificial Neural Network is a branch of artificial intelligence inspired by the biological brain. It is a structure of connected nodes (i.e., artificial neurons) that are arranged in layers. The links between nodes have weights associated with them depending on the amount of influence one node has on another. A feedforward ANN is an ANN without any cycles in the network and propagates incoming data in a forward direction only. A multilayer perceptron (MLP) is one type of feedforward ANNs with full connection of each layer to the next one. The MLP consists of one input layer, one or more hidden layers, and one output layer. Nodes in the input layer respond to data that is fed into the network, while output nodes produce network output values. Hidden nodes receive the weighted output from the previous layer’s nodes.

The most common concrete algorithm for learning MLP is the backpropagation algorithm. In the backpropagation algorithm, the computed output values are compared with the correct output values to compute the value of error function.Then, the error is fed back through the network.The derivatives of the error function with respect to the network weights are calculated. The weights of each connection are adjusted using gradient descent method such that the value of the error function decreases. This process is repeated until the network converges to a state with small error. Different training algorithms have been so far applied to train a backpropagation MLP. Each algorithm adjusts the ANN as follows:



**Machine Learning for Recommender systems**

if your problem is linear, you should go for

logistic regression (i.e. LR) or support vector machine (i.e.

SVM) and If your problem is non-linear, you should go for

k-nearest neighbor (i.e. k-NN), naive Bayes, decision tree or

random forest but from a business point of view, you would

rather use, logistic regression or naive Bayes when you want

to rank your predictions by their probability. For example, if

you want to rank your customers from the highest probability

that they buy a certain product, to the lowest probability.

Eventually that allows you to target your marketing

campaigns. And of course, for this type of business problem,

you should use Logistic Regression if your problem is linear,

and Naive Bayes if your problem is non-linear.

**Recommendations as a classification problem**

* **Regression - usage/interaction as a series of values (ex: probabilities)**
* **Classification - usage and other features are categorical (binned considering series ranges) - what we do is a multiclass classification**

subscriber demographic data such as age, occupation, gender,

income and call activities

**Text analytics can also be applied to CRM logs and/or social media to gauge customer sentiment and mine user opinions**

**User ratings:**

The interest of a user on an item is usually measured by a rating which can be obtained either

explicitly or implicitly.

* Numeric: when numerical values are assigned to products/services, for example, the five stars on the Amazon website
* Ordinal: when the user is prompted to select a term that best indicates his/her opinion on an item, such as “I agree”, “I am neutral” and “I disagree”
* Binary: when the user simply decides if an item is good or bad
* Unary: this kind of ratings was popularized by Facebook where users can mark his/her interest in a post or photo by clicking a button “Like”

there are two main methods of **collaborative filtering** , the nearest neighbor methods and the latent factor methods.

There are several kinds of **CF methods**, among them the most popular approaches are user-based CF and item-base CF. A user-based CF method is to use the ratings of users those are most similar to the target user for predicting the ratings of unrated items. On the other hand, item-based CF method uses the similarities of items for predicting ratings.

Literature shows that the current trend of recommender system is to combine two or more techniques together for improving the accuracy of recommendation or overcome the limitations of single recommender algorithm, and the combination of user-based CF and item-base CF may achieve a good performance in a big-user-set and big-item-set environment.

Steps:

1. Generate a User-item Rating Table: Each user

is represented by a set of item-rating pairs and the

summary of all those pairs can be collected into a

user-item rating matrix.

1. Calculate Item Similarity: This step measures

the similarities between any two items. Pearson

correlation is selected for this step which measures

the similarity between two items by calculating the

linear correlation between the two vectors.

1. Item Neighbours Selection: In most CF

methods, a number of neighbours will be selected

when predicting ratings. In the TCPRS, we used the

top-N technique for neighbour selection.

1. Predict Empty Ratings using Item-based CF: In

this step, all the unrated ratings can be calculated

using item-based CF method and all the empty cells

in the user-item rating table will be filled.

1. Calculate User Similarity: Beside from

predicting the ratings based on the similarities of

items, we can also predict the ratings by analysis the

similarities between users. We also use the Pearson

correlation algorithm for calculating the user

Similarity.

1. Select Top-N Similar Users : Similar as step 3,

we need to select a number of neighbour users for

predicting ratings. The Top-N technique is used in

the TCPRS system.

1. Final Recommendation Generation: The final

step of the algorithm is to predict the ratings of

every unrated telecomm product/services for the

target users using user-based CF. The new predicted

ratings will replace the ratings predicted in Step 4,

and be regarded as the final results.

**\*\* Nearest neighbor**

The nearest neighbor methods are based on the principle that users who have preferred similar items in the past tend to prefer similar items in the future.

**\*\* Latent factor**

**Approaches to resolve cold-start problem in NN CF:**

**ItemUserAverage**: Like ItemAverageRecommender, except that estimated

preferences are adjusted for the Users' average preference value. For

example, say user X has not rated item Y. Item Y's average preference value

is 3.5. User X's average preference value is 4.2, and the average over all

preference values is 4.0. User X prefers items 0.2 higher on average, so, the

estimated preference for user X, item Y is 3.5 + 0.2 = 3.7.

**TreeClustering**: A Recommender that clusters Users, then determines the

clusters' top recommendations. This implementation builds clusters by

repeatedly merging clusters until only a certain number remain, meaning that

each cluster is sort of a tree of other clusters.

**Opinion Mining**

With the growth of social networks, more and more users can openly discuss their impressions

and experiences on a variety of products, items, and services. This means a significant increase in user-generated content in the form of reviews, blogs, discussion forums, social networks, etc. Among this content, reviews represent rich sources of data and they are very useful for marketing intelligence, social psychology and other areas that are interested in mining opinions, views, sentiments and attitudes

**Analysing sentiment of reviews:**

Frequency-based: an aspect can be expressed by a noun, adjective, verb or adverb, but studies

show that from 60 to 70% of explicit aspects are nouns [69]. Aspects tend to be frequent nouns

since, in commentaries, people are generally more likely to talk about the relevant aspects.

Based on syntactic relations: there are usually many syntactic relationships between the

expressions of sentiment and the opinion targets. Such relationships are possible to be explored

when words and phrases of sentiment are known.

Through supervised learning: in general, methods for identifying aspects are based on sequential labeling. The most commonly used methods are: **Conditional Random Field (CRF)** and **Hidden Markov Model (HMM)** .

**Challenges in telcom Big data:**

Variety:

Today, the proliferation of applications enabled by the Web, mobile networks, GPS, and social media has forever changed the horizon.

The numerous data points created and made avail- able by these applications have resulted in virtually a “data rainforest” with highly diverse sources of structured (tab- ular), semi-structured (objects, log records), unstructured (free text), and streaming data

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Informations Variety in Telcom Domain**

The information may include, e.g. what websites you visited, how much time you spent talking on the phone, watching video, and on OTT (over-the-top) apps such as Skye,

WhatsApp, and Facebook. By pairing these usage data with the network KPIs the operators are able to gain more insights into users’ QoE (Quality of Experience) on different services.

**\*\*\*\*\*\*\* Big data/ Capacity**

Telecom data has grown exponentially since the age of broadband and 3G. This trend

will continue as the networks evolve (optical, 4G/LTE, 5G), allowing users to access, contribute, and share ever increasing contents on the Internet.

**\*\*\*\*\* technology Availability**

Many open sources have either a free community version or an enterprise version with moderate service/license fee. The Apache projects [13] are one of the most prominent examples.

**A suggested telcom framework:**

framework contains three horizontal layers – resource, service, and customer, spanning across two vertical perspectives – infrastructure & product and operations

**\*\*Note:**

**Resource layer:** activities related to network build-out, planning, and monitoring. Operators constantly monitor performance of the networks (devices such as routers, switches, base stations, etc.) in order to assure smooth operation.

Data collected at this layer:

* alarms generated by the network devices
* key performance indicators (KPIs) such as
  + packet loss ratio,
  + latency,
  + traffic load, etc.

\* The datasets support :

* network planning,
* capacity management,
* fault management.

**service layer**: includes activities related to provisioning of user services (voice, data, and video). It also supports **proactive monitoring** and **reactive diagnostics** required by

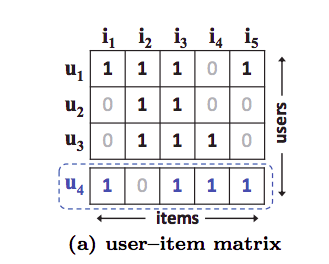
service-level agreements – a contractual agreement between the operator and the users on the performance and availability of the subscribed services.

History logs from service provisioning can be used to improve the process,

Usage pattern data can be mined to detect frauds or monetized by selling to companies that are interested in reaching out to potential customers.

**customer layer:** the main task is Customer Relationship Management (CRM), which handles user inquires, orders, trouble tickets, and assure user satisfaction

**Cosine similarity : problem - cold start**



Used in : matrix factorization

Suppose, this is our item-user interaction matrix. We will use the cosine function to get the similarity between the user Latent vectors.

S{i,j}= Cosine(i,j)

S denotes similarity between user i and j.

So, we get,

S{23} > S{12} > S{13}.

Now, if we arrange them according to their similarities we obtain:

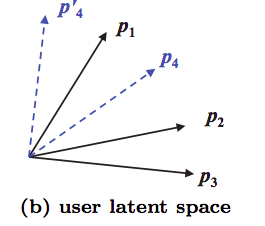
Now, a new user comes in U4.

On recalculating, we will get.

S{41} > S{43} > S{42}.

If we observe closely there is a confusion U4 is most similar to P1, then P3 and finally P2.

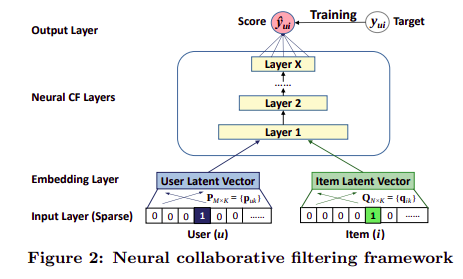
Let’s see another vector representation:



If by any chance u4 is placed p4, it will be similar to u2 and give the wrong recommendations as it will give u2 more weightage than u3 for prediction of u4.

**This drawback is answered by deep neural network**

Architecture of **Neural Collaborative Filtering:**



Working: The model takes in two sparse vectors, one representing the user and the other represents items. The item vector has 1 at an index means the user has interacted with the item corresponding to the index. So, elaborately,

User vector=[ 0,0,1 ………..0] with m elements, means this vector represents the 3 rd user out of m.

Item vector=[0,1,0,1,1,0…..1] with n elements, means the user interacted with those items out of n items.

Basically both items and users are one-hot encoded.

The next layers are the embedding layers that obtain the dense or latent vectors for the sparse inputs, from the input layer. Now, the obtained latent vectors are fed into the multi-layer neural architecture, to map the latent vectors to the predicted probability scores. The layers are responsible to find the complex user-item relations from the data. The output layer produces the predicted score Y\_pred(ui), i.e, how much is the probability that the user u will interact with the item i.

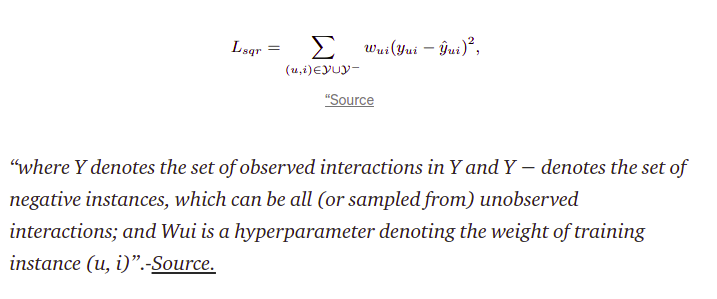
For training the model, the authors have used pointwise loss function, to minimize the difference between the target value Y(ui) and the corresponding predicted value.

The NCF model’s Prediction equation is given by:

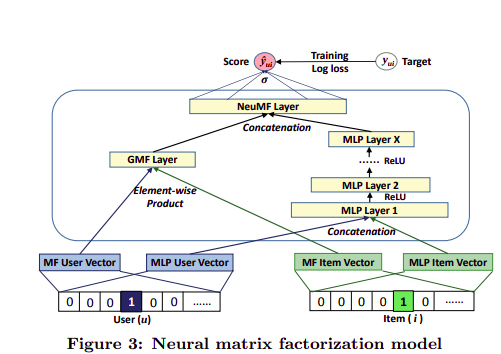


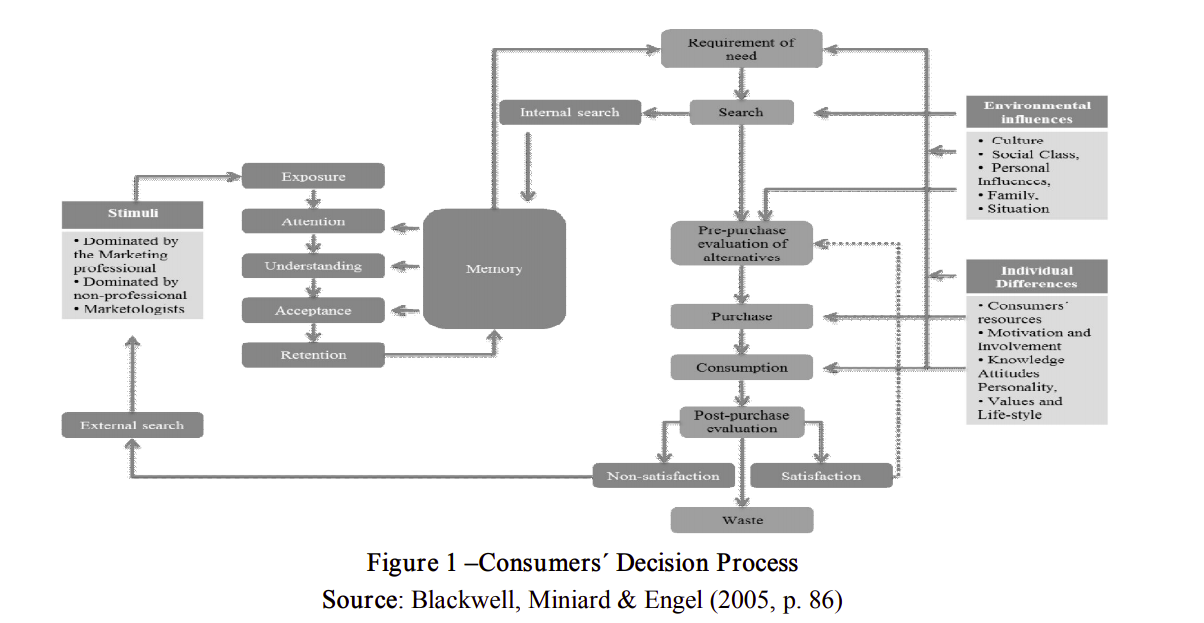
“where P ∈ R(M×K) and Q ∈ R (N×K), denoting the latent factor matrix for users and items, respectively and Θf denotes the model parameters of the interaction function f”.

The NCF uses a pointwise loss to train model parameters. So, it trains using a mean squared loss. The loss is described as:



Neural matrix factorization model:





Readings

## 1 Context aware

Paper 1: Context-Aware Recommender Systems

*Gediminas Adomavicius, Bamshad Mobasher,*

*Francesco Ricci, and Alex Tuzhilin*

Context - current situation, a state

Ex:

For example, when a user is buying books, the preferences the user expresses in one context, such as “books for my children,” may be of no predictive value when the user seeks recommendations in a different context, such as “work-related books.”

Traditional hybrid recommenders: based on content-based and collaborative filtering, tend to use fairly simple user models.

Recommend on similar user preferences, user as a vector of item ratings

Context: representational/ interactional

* Representational: Observable attributes that are known a priori. The structure does not change over time.
* Interactional: The scope of the features is defined dynamically. Cyclical relationship between context and activity (which makes the context change).

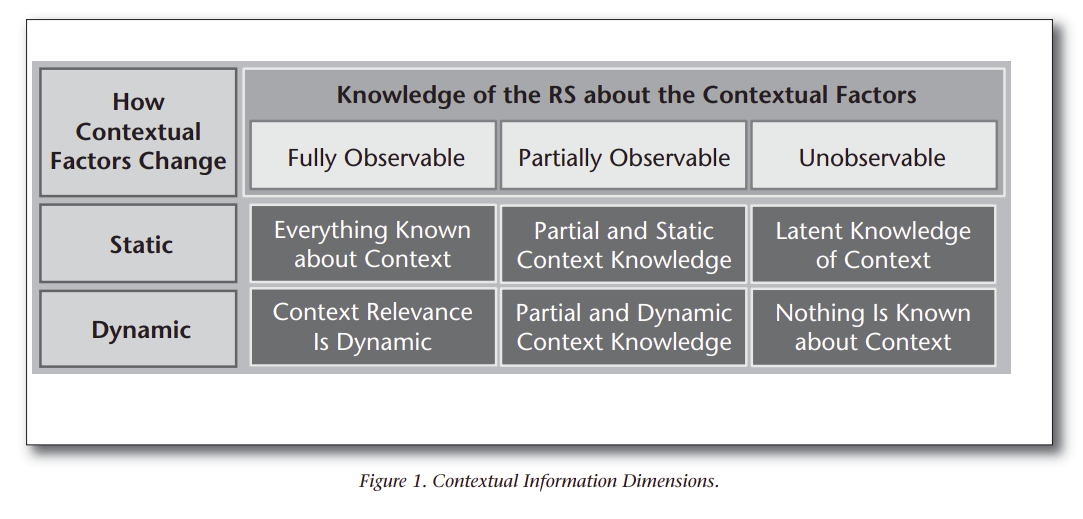
Goal: how context can be defined and used in recommender systems in order to create more intelligent and useful recommendations

Contextual factors: time/location/ purchasing purpose

Knowledge of contextual factors: Fully observable/ partially/ unobservable

If unobservable:

* unobservable context is modeled using latent variables.
* the recommender system may build a latent predictive model, such as hierarchical linear or hidden Markov models,



Contextual factors: change over time -> static/ dynamic

Dynamic: some inputs to recommendations, may not be valid within the time

If structure of a contextual factor is static:

* unobserved contextual information can be learned using some of the machine-learning methods, such as matrix factorization (Koren 2008), probabilistic latent semantic analysis (PLSA), or hierarchical linear models (HLMs).

Note:

Dynamic context may change based on passive observations or explicit user feedback.

**Ex:**

**Conversational systems: ex: chat bots**

**user feedback is used to iteratively refine the user profile (or the initial user query) resulting in more appropriate recommendations.**

**In a contextaware conversational system (for example, Baltrunas et al. [2011]), the user feedback may also be used to iteratively modify contextual factors and not just user profiles.**

For example, in the course of conversation with a user, a restaurant recommender may determine that the user is on a romantic date. This observation, in turn, may result in filtering out restaurants that tend to be noisy or without an adequate wine selection.

**Problems with representational sys**

Problem with representational (fully/static) - these context variables need to be defined at the system design stage( ex. Get user data on a purchase), or need to determined as part of the data collection - not feasible to achieve

It is often difficult to determine a priori what all relevant contextual factors are.

For example, in a restaurant recommendation domain, it may be difficult to determine whether the user’s attire would be relevant to his or her choice of restaurant.

**Traditionally, the recommendation problem has been viewed as a prediction problem in which, given a user profile and a target item, the recommender system’s task is to predict that user’s rating for that item, reflecting the degree of user’s preference for that item.**

Traditional:

Users x Items -> Ratings

Context-Aware:

User, item , context ,rating

where each specific record includes not only how much a given user liked a specific item, but also the contextual information in which the item was consumed by this user (for example, context = Saturday).

**\*Contextual pre-filtering:**

We use context to select appropriate 2D recommender (user x rating)

An example of a contextual data filter for a movie recommender system would be: if a person wants to see a movie on Saturday, only the Saturday rating data is used to recommend movies.

Note:

**The exact context sometimes can be too narrow.** Consider, for example, the context of watching a movie with a girlfriend in a movie theater on Saturday or, more formally,  
 c = (Girlfriend, Theater, Saturday).

Using this exact context as a data-filtering query may be problematic, because certain aspects of the overly specific context may not be significant,

**\*Contextual post-filtering:**

Not account context data at input data, when generating recs.After generated, context is used to

(1) filtering out recommendations that are irrelevant in a given context, or

(2) adjusting the ranking of recommendations in the list.

3 approaches in brief:

* Prefiltering: Context is used to select some set of data and then predict like a traditional recommender system. One way of doing that is splitting items or users by context as if they were different items or users.
* Postfiltering: Ratings are predicted and then the results are filtered using the context. This can be implemented by ordering the results depending on the context or by just filtering the results.
* Modeling: The context is used right in the model. It is more complex and could be implemented by multiple machine learning models like SVM, matrix factorization or a markov chain.
* Physical context: representing the time, position, and activity of the user, but also the weather, light, and temperature when the recommendation is supposed to be used.
* Social context: representing the presence and role of other people (either using or not using the application) around the user, and whether the user is alone or in a group when using the application.
* Interaction media context: describing the device used to access the system (for example, a mobile phone or a kiosk) as well as the type of media that are browsed and personalized. The latter can be ordinary text, music, images, movies, or queries made to the recommender system.
* Modal context: representing the current state of mind of the user, the user’s goals, mood, experience, and cognitive capabilities.

## Paper 2: Introduction to recommender Systems

Demographic filtering, content-based filtering, collaborative filtering and hybrid methods are the main four methods

Collaborative - user ratings - KNN widely used

* + based on two main approaches: user to user and item to item.
  + kNN algorithm first tries to determine the k neighborhood for the user; then, it aggregates users based on their ratings and finally predicts based on the aggregated information.

Content-based: content of the item that user liked in past

Demographic - get user’s common attributes, suggest recs according to them

Why not one kind ? why need hybrid?

1. Cold start is one of the main challenges that almost all recommender systems face when the initial ratings or any knowledge about user experience is not sufficient
2. Cold start: new community/ new user / new item
3. In the new community problem, the RS suffers from a lack of sufficient data when it is initialized for a new community.

Getting active users feedback (ratings etc..) into a recommender system:

Methods: explicit/ implicit

Explicit: ex: User’s rating for an item

**Implicit: generate a rating from users behaviour data (in SLT case)**

ex:if a user tries action movies frequently, the implicit information implies that the user’s rating on action movies could be high.

**----> used in SLT case (to generate ratings)**

to design demographic filteringsystems, we need some information about users to categorize them into groups.

Content-based methods make recommendations based on the description of the items. Nowadays, it is combined with other methods and use more information about items and users.

Ex: to recommend movies to users. We may assume that the movie description has been already extracted. If the movie is an action film and a user liked it, the recommender system will recommend another action movie to the user.

3 parts:

preprocessing on items with a **content analyzer**,

then a **profile learner** learns about users.

Finally, the **filtering component** finds a set of appropriate recommendations.

• Content Analyzer — For any decision making problem, the raw data should be pre-processed to extract featured information. Here, the output of this pre-processing part is the structured relevant information. The content analyzer prepares information for the next step.

• Profile Learner — This module is specifically designed for the user side. It receives the pre-processed information from the content analyzer and generalizes them to construct the user preferences. The generalization step models the user interest based on the user’s past ratings for items.

**Content based: techniques:**

1. Keyword based vector space model

both user profiles and items are represented by vectors of weighted terms

Currently: semantic aware methods

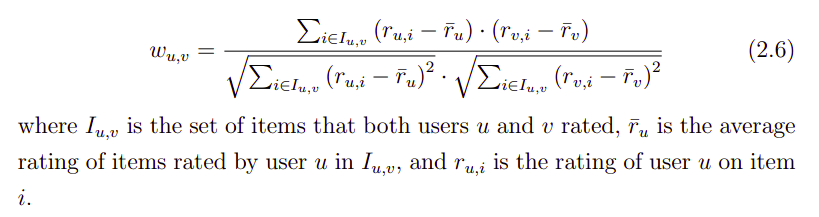
**Collaborative Techniques**

Now, if two users named “Marcos” and “Diego” like a movie titled A, and later Marcos watches another movie titled B and likes it, then we can recommend this movie to Diego. This approach is adopted from the collaborative filtering method.

1. Cosine similarities
2. KNN - Neighbourhood approach
3. **latent factor model**, (matrix factorization/ ALS) : transforms both items and users to the same latent factor space.

Divided into two:

1. Memory based
   1. User-item rating matrix - can easily be adapted to use all the ratings before the filtering process
   2. Train: 100% pass data -> predict depend on them
   3. Similarity measure - **Pearson Correlation measure (with Nearest Neighbors)**
      1. reveals the information on how much two variables are linearly related to each other

****

1. Model-based
   1. Ex: a neural network based approach: , generates a model that learns from the information of user-item ratings and recommends new items.
   2. Train set + Evaluation set -> model trained and predict future data

**Hybrid Techniques:**

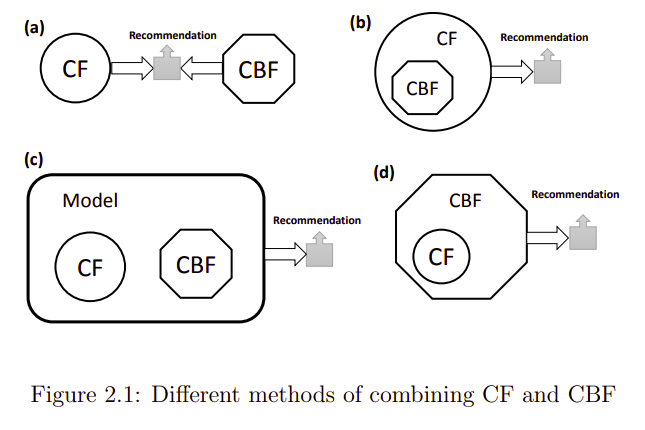
A hybrid filtering method may use a combination of collaborative filtering with demographic filtering or collaborative filtering with content-based filtering to have boosted results.

Ex:

Balabanovic et al. [5] created a recommender system named Fab which extracts user profile 15 of interest on web pages by content filtering techniques and uses that information for collaborative filtering.

Content based -> multi-level collaborative filtering

4 different methods:



Mostly CBF -> CF

But in some cases, CF -> CBF (user ratings -> categorize users better)

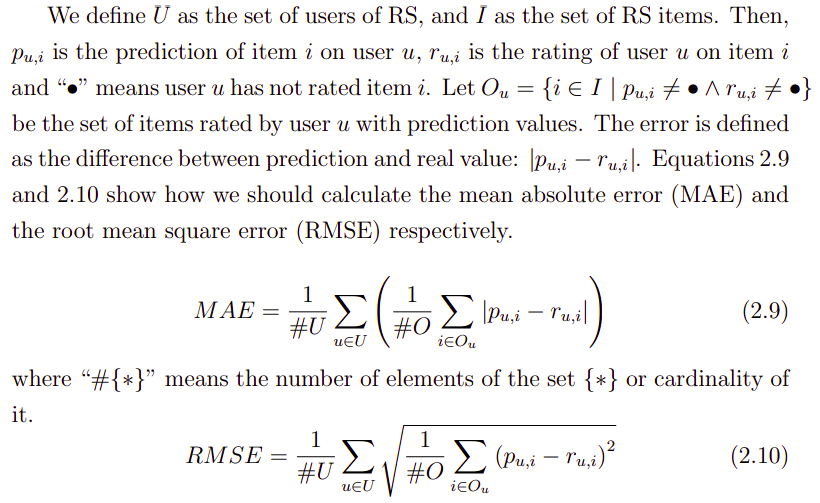
**How recommender systems are evaluated**

Quality of Predication

MAE (Mean Absolute error)

RSME (Root Mean square error)

Coverage



Mostly : **having a reduced set of items is more important that having one item recommended.**

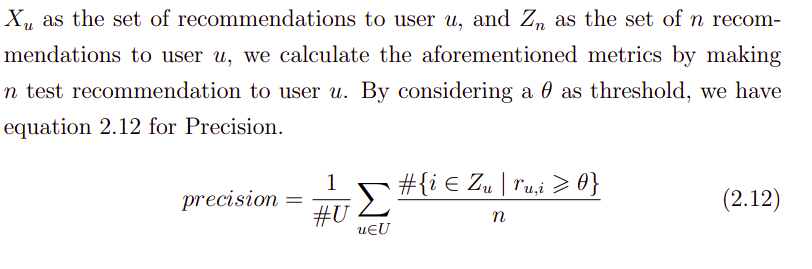
Evaluation quality of a set of recommenders:

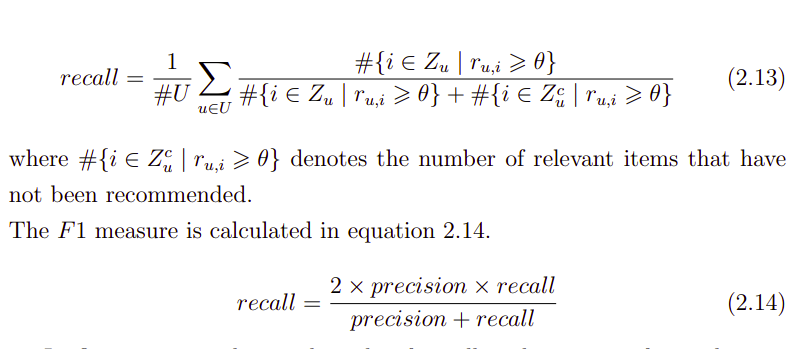
Precision

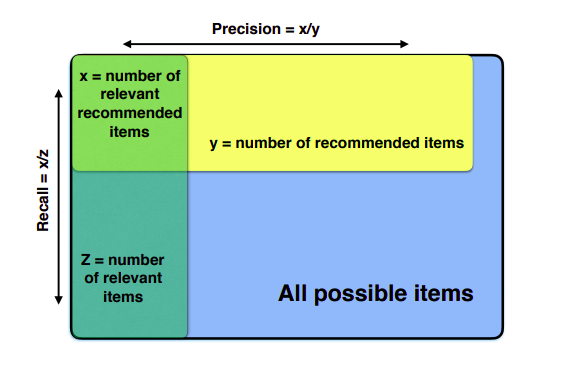
Recall

F1

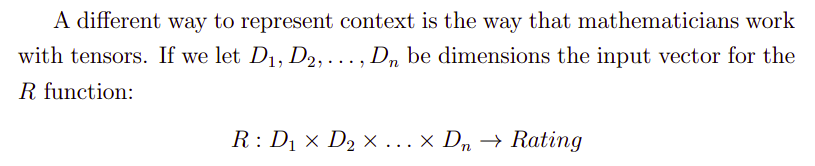
1. Precision indicates the rate of relevant recommended items to all of the recommended items.
2. Recall is about the rate of relevant recommended items to all of the relevant items
3. F1 is a combination of precision and recall.



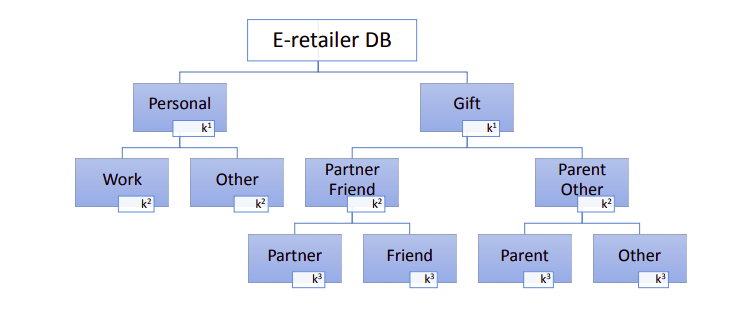




**Context aware recommenders:**



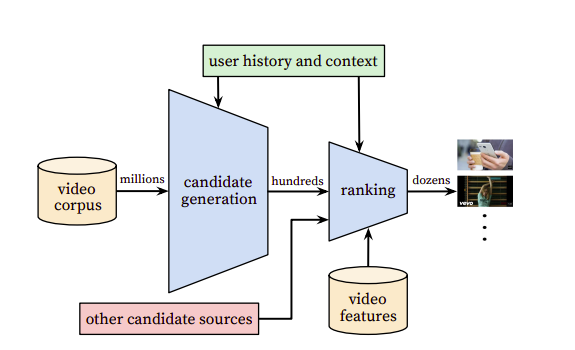
Ex: contextual info in a hierarchy



**Deep learning for recommender systems:**

Youtube Deep Learning model:

Historical user behavior on YouTube is inherently difficult to predict due to **sparsity and a variety** of unobservable external factors.



**Other papers:**

content-based setting, Burges et al. used deep neural networks for music recommendation [21].

collaborative filtering is formulated as a deep neural network

2 neural networks :

Candidate generation

Ranking

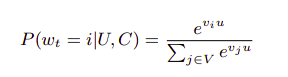
The candidate generation network takes events from the user’s YouTube activity history as input and retrieves a small subset (hundreds) of videos from a large corpus.

**Note:**

**recommendation as extreme multiclass classification where the prediction problem becomes accurately classifying a specific video watch wt at time t among millions**

**of videos i (classes) from a corpus V based on a user U and**

**context C,**



**implicit feedback** [16] of watches to train the model, where a user completing a video is a positive example. This choice is based on the orders of magnitude more implicit user history available,

**Implicit feedback** is data we gather from the users behaviour, with no ratings or specific actions is needed.

It could be what items a user purchased, browsing history, search patterns, how many times they played a song or watched a movie, how long they've spent reading a specific article etc.

Sampling - To efficiently train such a model with millions of classes, we rely on a technique to sample negative classes from the background distribution

we learn high dimensional embeddings for each video in a fixed vocabulary and feed these embeddings into a feedforward neural network.

**Collaborative model:**

**Extension of matrix factorization using deep NN**

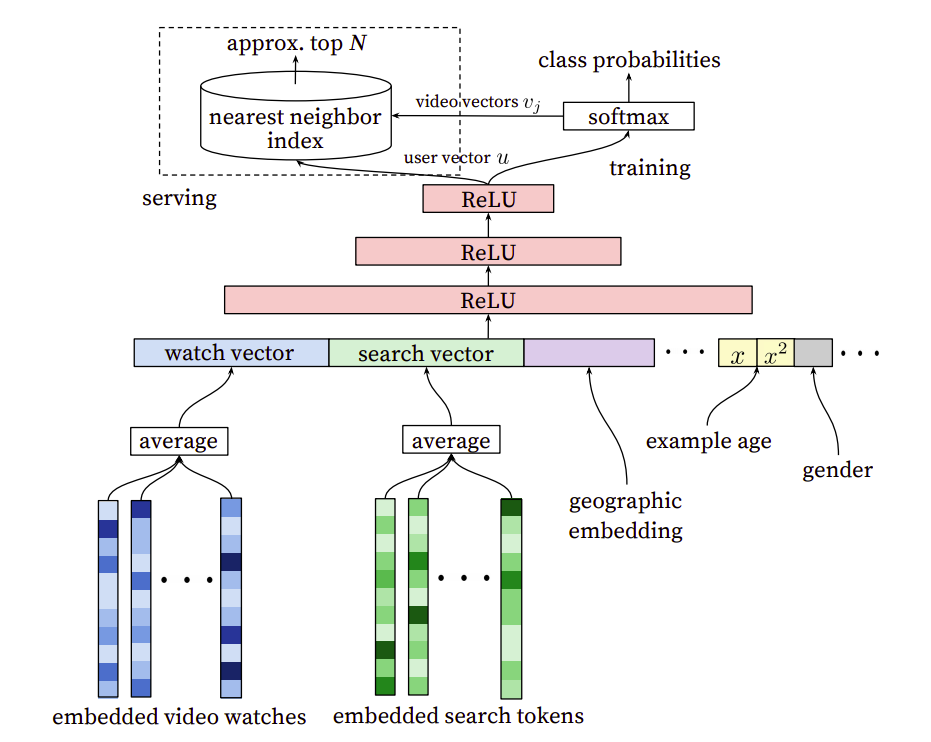
A key advantage of using deep neural networks as a generalization of matrix factorization is that arbitrary continuous and categorical features can be easily added to the model. Search history is treated similarly to watch history - each query is tokenized into unigrams and bigrams and each token is embedded.

Why deep Learning?

* Machine learning systems often exhibit an implicit bias towards the past because they are trained to predict future behavior from historical examples. T
* Multi class classification
* Adding more categorical feature into extended matrix factorization

It is important to emphasize that recommendation often involves solving a surrogate problem and transferring the result to a particular context.

Candidate generation model:



Feature engineering:

embedded sparse features concatenated with dense features. -> features to train the model

**we use embeddings to map sparse categorical features to dense representations suitable for neural networks.**

**Very large cardinality ID spaces (e.g. video IDs or search query terms) are truncated by including only the top N after sorting based on their frequency in clicked impressions.**

Importantly, categorical features in the same ID space also share underlying embeddings. For example, there exists a single global embedding of video IDs that many distinct features use (video ID of the impression, last video ID watched by the user, video ID that “seeded” the recommendation, etc.).

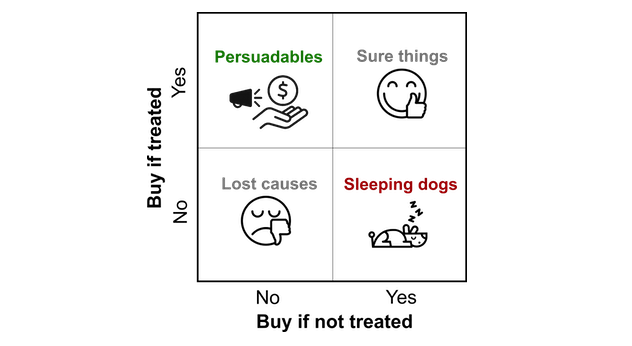
**Uplift Modeling**

probability of purchasing (if they receive the offer) minus the probability of purchasing (if they *do not* receive the offer).

**A similar situation occurs in retention. Consider a retention campaign where customers are selected to be contacted by a retention team to persuade them to continue their subscription. The goal is to only contact customers who can be persuaded to stay.**

**We want to avoid contacting customers who always intended to stay or leave, regardless of being contacted.**

**More importantly, we want to avoid contacting customers who may actually be prompted to leave because we have contacted them**

****

**The premise behind uplift modeling is that it only makes sense to send marketing interventions to persuadables, because they are the only group where targeting them gives a better outcome than not targeting them.**

Modeling approaches:

**causal conditional inference forest (CCIF)**

which is based on binary decision trees. At each node in the tree, it tests whether there is any interaction effect between the treatment and the features. If any of them show an effect, then a split is made using the feature showing the strongest effect.

**Transformed Outcome Tree**

This transformed outcome is constructed such that its average value corresponds to the average uplift. The benefit of this approach is that once the simple transformation is done, the uplift problem becomes a binary classification problem, and any of the standard machine learning algorithms can be applied.

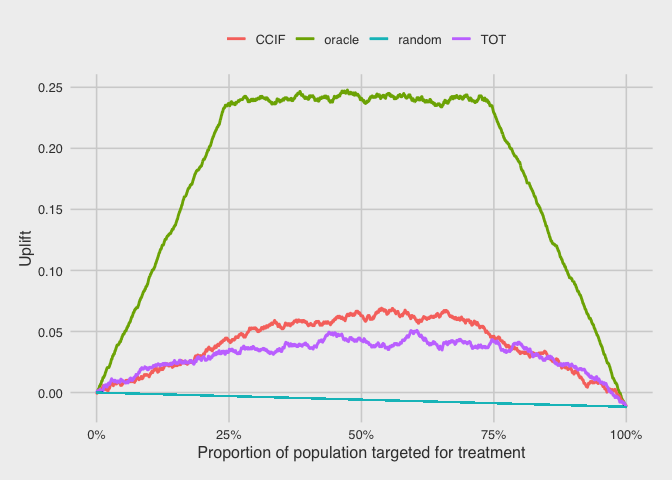
Python library: pylift

**Uplift evaluation: by decile**

A simple way to evaluate uplift models is to compare their predicted uplift with their actual uplift.

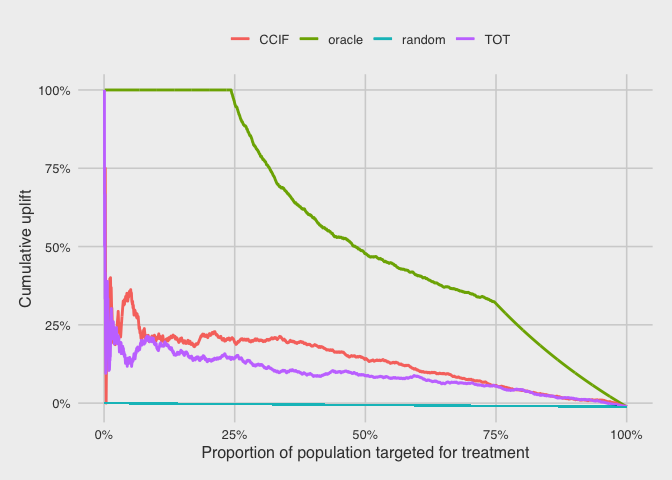
since actual uplift cannot be calculated for each individual customer (i.e. they cannot receive and not receive the offer simultaneously), we calculate the actual uplift over a group of customers. To do this, we rank the customers in the test dataset according to their predicted uplift, and partition them into deciles.

Qini curve:



it plots the cumulative uplift across the population. Here, we rank the customers by their predicted uplift on the horizontal axis, and the vertical axis plots the cumulative number of purchases in the treatment group (scaled by the total treatment size) minus the cumulative number of purchases in control (scaled by the total control size).

Cumulative uplift:

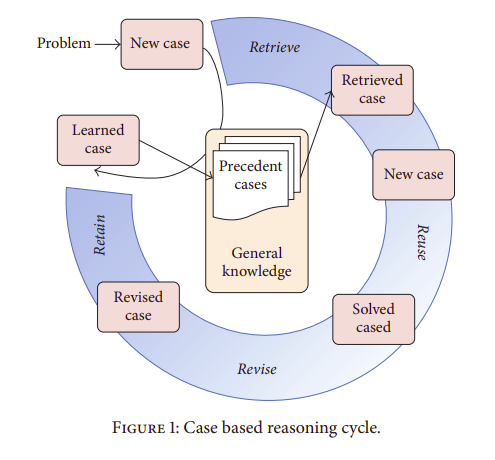


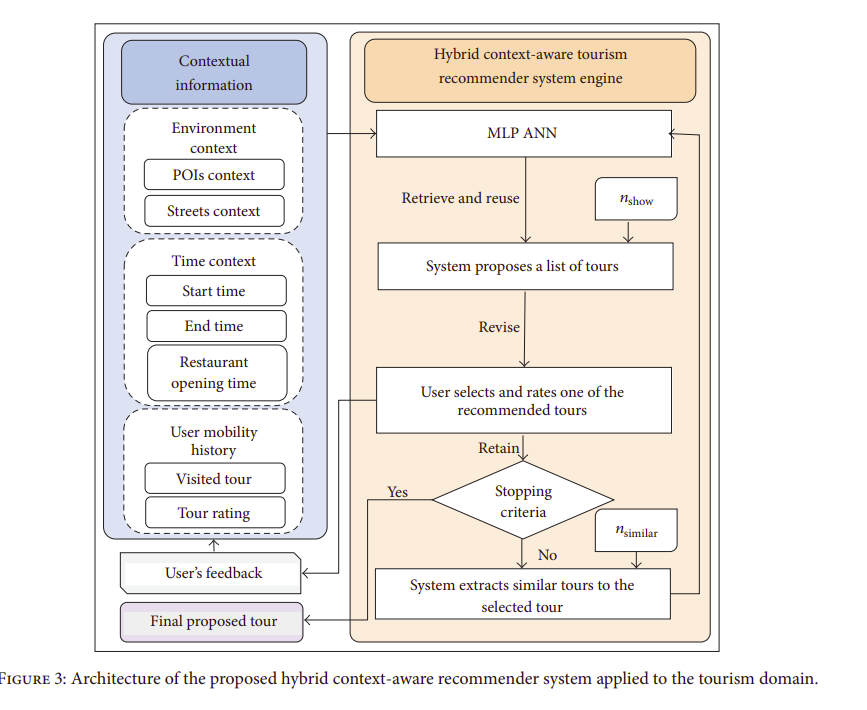
Another way to compare the models is to plot their cumulative uplifts. Again, we rank the customers by their predicted uplift on the horizontal axis. The vertical axis plots the cumulative number of purchases in treatment (scaled by the cumulative treatment size) minus the cumulative number of purchases in control (scaled by the cumulative control size).

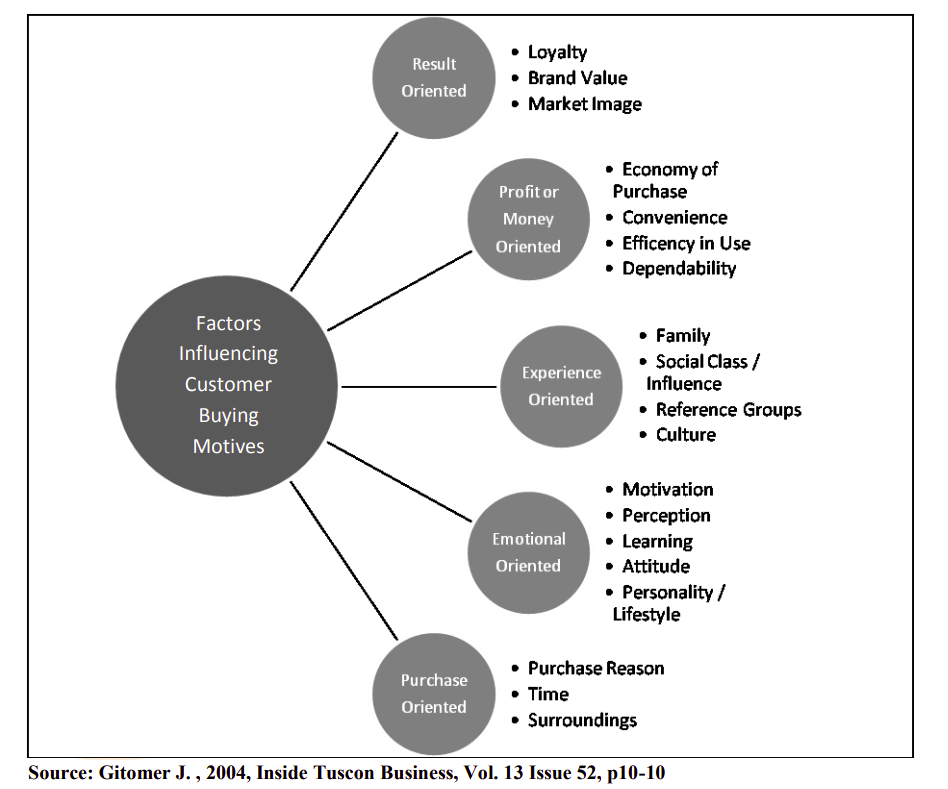
**Paper: Context aware deep learning recommender for travel**

contextual parameters such as user’s own mobility, previous history, and the timing of a recommendation should be considered as important contextual information.

Case based reasoning:







**Paper 01 : Factors affecting brand switching behaviour (Experience from Sri Lanka telecom Industry) 2020**

Findings:

**service quality and price** were the major factors that entices customers to switch their telecommunication brand

Reasons for brand switching:

**accessibility, value, consistency, price, promotions and receptiveness.**

newer technology, better communication and considerate services are the hallmark of quality in the telecommunication industry (Makwana et al., 2014).

main problem is that **service providers are not in complete control of consumers’ switching decisions.** It is up to the consumer to choose the service provider who will generally be located in close proximity

network market in Sri Lanka is **oligopolistic**, and in the current market the **number of service providers is very limited but yet the competition is very high**.

\*\* market share : analysis

Quality

Service quality may be described as a factor in which the actual offerings obtained by customers are compared to their expectations regarding it.

Telcom services: quality -> functional quality, technical measures

**\*\* Service quality depends upon assurance, network quality, value-added services and responsiveness**

**→ recommending value added services: will affect positively**

Price

If the service provider gives excellent service and charges a fair price, that is likely to satisfy the customer

Price does not only mean the service cost, as it is also related to the value as perceived by the customer

**\*\* A research conducted in Pakistan has observed that brand switching by cellular phone customers has frequently been shaped by price rationality**

Brand Image

, it is important to build up a good brand image as that is the only way to earn the public’s confidence in a company’s products and services

Customer satisfaction

defined customer satisfaction as the level of contentment

of a consumer, based upon his/her expectations in respect of a product or service of a brand.

From the price, the customer can feel that he/she is getting his money’s worth, while from the quality of service, the customer can feel satisfied overall with the service provider

there are two types of customer satisfaction, which are

transaction-specific satisfaction and

cumulative satisfaction.

Customer’s valuation of service or product after choosing and utilizing it is transaction-specific satisfaction

Cumulative satisfaction is a long-term process as it includes the assessment of

service or product over a longer period, depending upon its usage and quantity

and frequency of purchase

**Hypothesis:**

Factors affecting customers buying behavior:

Service Quality

Fair price

Brand image (brand loyalty)

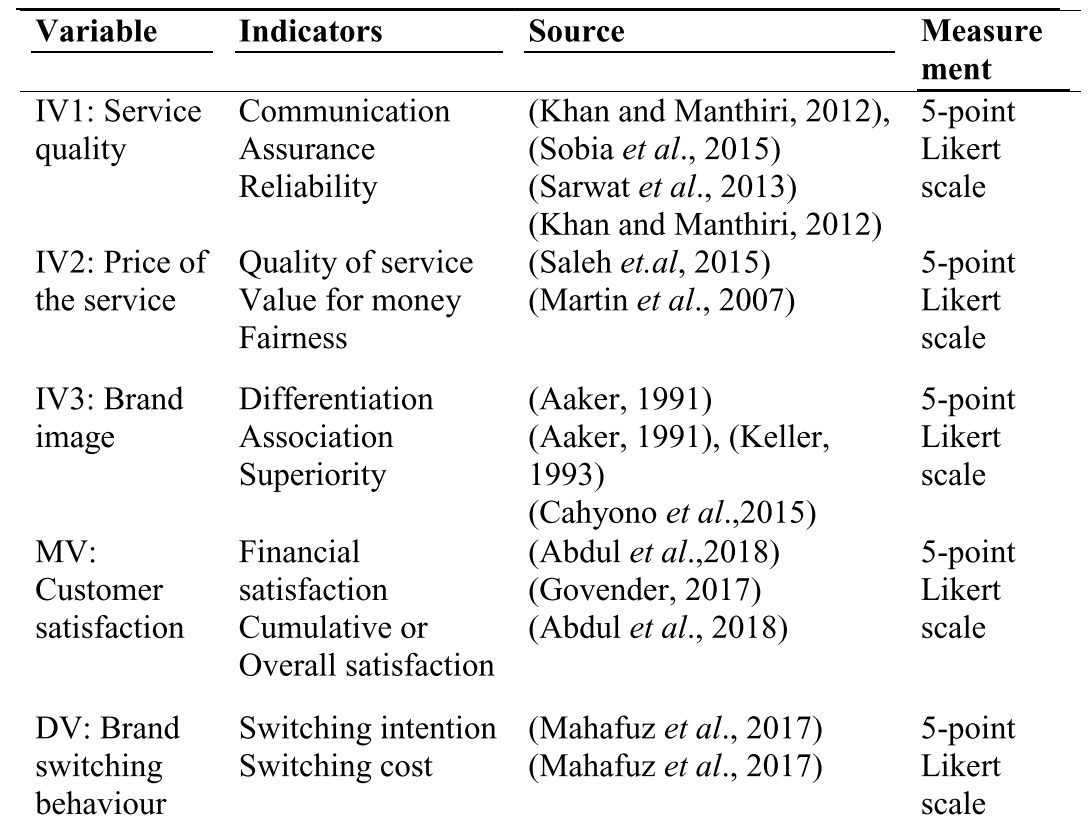
Customer satisfaction (trnsaction-specific, cumulative)

When a telcom customer tend to churn? (switch brands)

Economical downfall of a company

When consumers arent brand loyal

**\*\* switching cost?**



**Paper 2: FACTORS INFLUENCING CUSTOMER BUYING MOTIVES; WITH SPECIAL REFERENCE TO SRI LANKA TELECOM BROADBAND AT MICRO LEVEL (2012)**

The research reveals that the emotional factor should be given top most priority at each respective marketing attempt with the required attention to other factors accordingly.

Though the massive customer awareness programs are coupled along with its products and services, the fixed line market faces intense competition from the mobile service providers while experiencing low response for its traditional core businesses.

Customer awareness programs:

pre-product launch,

after commercialization and

at the time of value added services are launched,

**\*Problems:**

Gradual decline in the Broadband market share

customer purchases are influenced strongly by cultural, social, personal, and psychological characteristics.

classification of buying motives by dividing them in to two extremes, those which are operational and those which are sociopsychological. He further explains that the operational buying motives as those reasons for the purchase that are directly related to the anticipated performance of the product and sociopsychological motives as reasons for the purchase that are indirectly related to the anticipated performance of the product and directly related to the consumers social and psychological interpretation of the product

A buyers’ decision also influenced by personal characteristics such as age, life-cycle stage, occupation, economic situation, life style, personality and self concept. Psychological factors including motivation, perception, learning, beliefs and attitudes are also should be taken as key concerns.

Data collection methods:

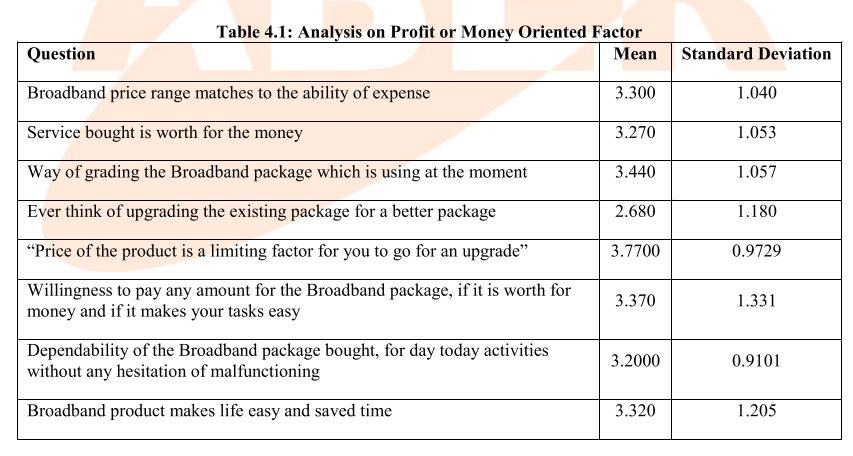
An in-depth discussion carried out with the SLT internal product development and marketing groups as well as a comprehensive survey was carried out based on the targeted customer segments in the country.

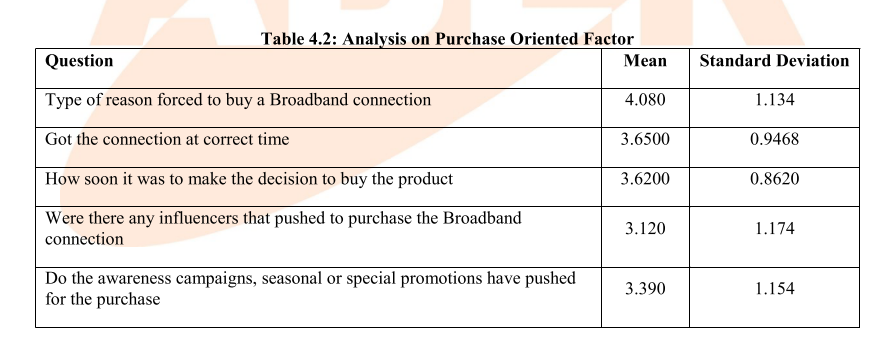
For gathering the customer information the SLT directory was the selected resource of

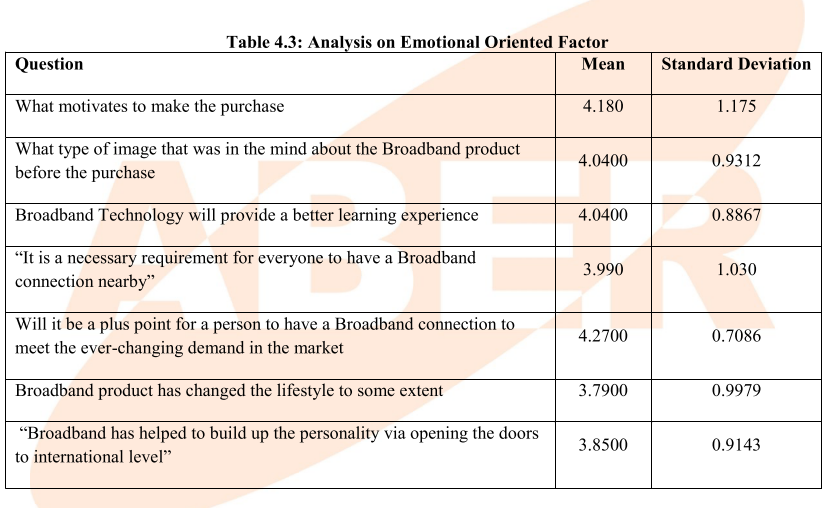
data. While selecting the sample, judgmental sampling has also been merged into the purposive sampling method where necessary. From the areas which were covered by the broadband product, the targeted segments have been selected as it contains different interest groups. Business customer (Micro level) and residential customer (Professional / None Professional) segments were considered as major two categories.

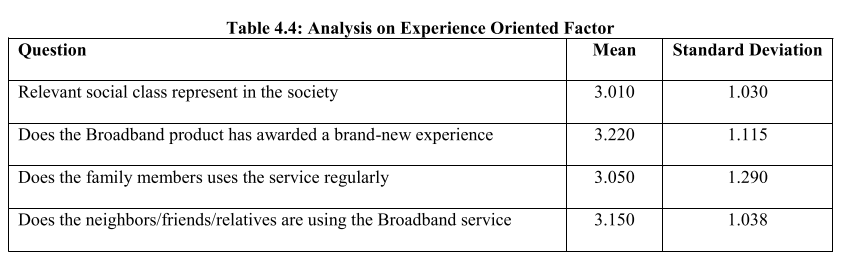
Therefore, each and every measurement variable has been evaluated through five point lickert scale system (1 to 5, where 5 = “Strongly positive position” and 1 = “Strongly negative position)

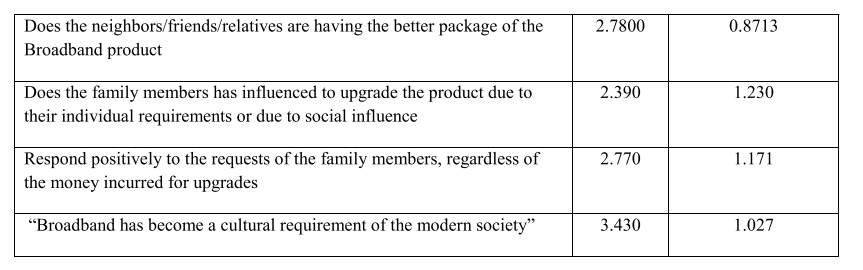
Paper analyse on - price/money, purchase idea, and emotion and expereince based buying decisions

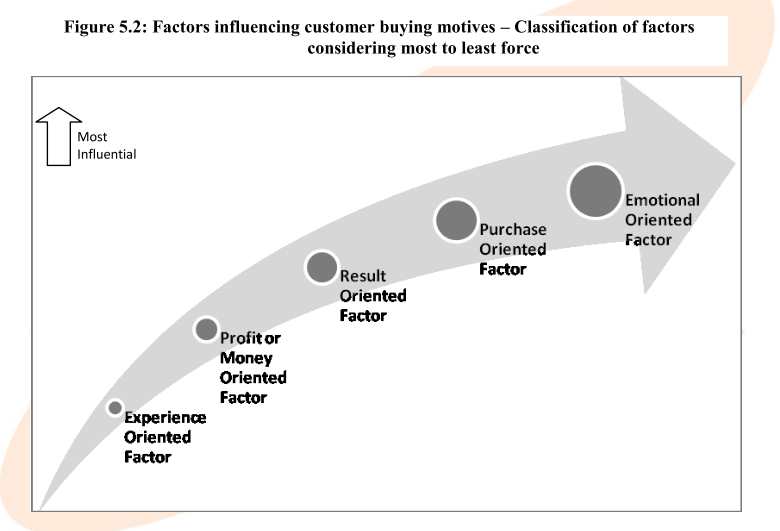








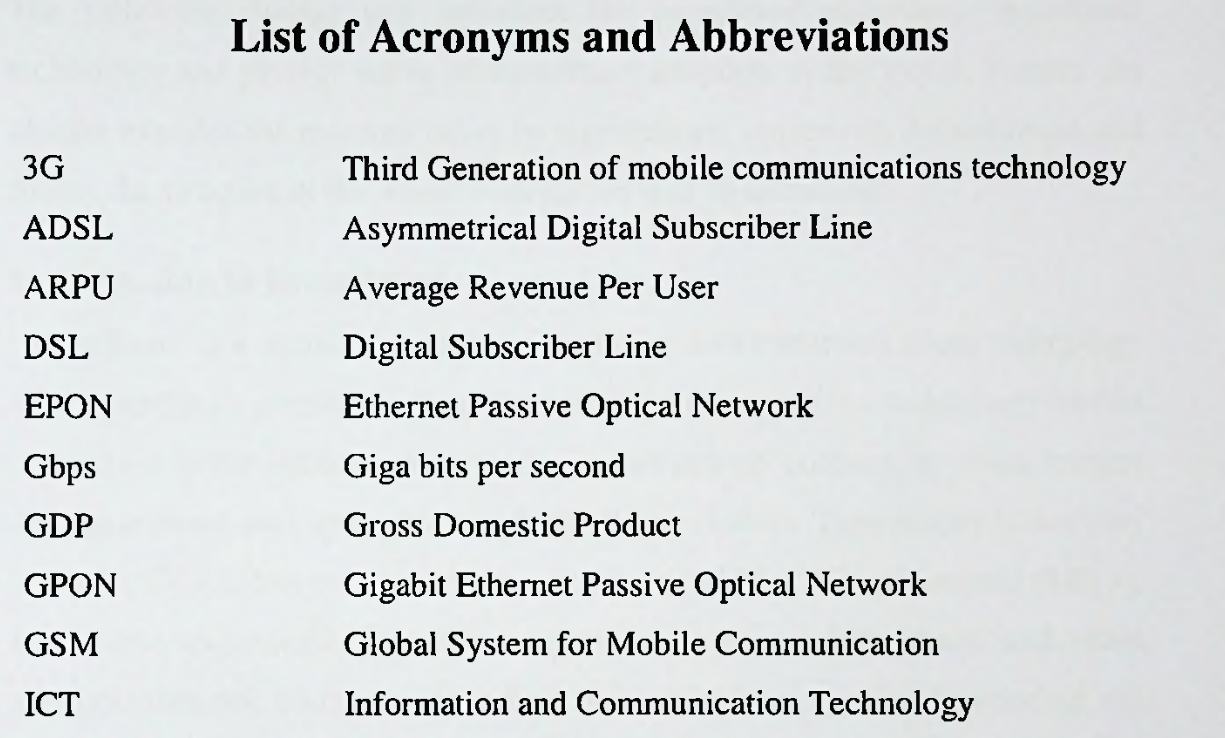




SLT should concern more on grabbing the hearts and minds of the existing as well as potential customer segments by taking the opportunity to the maximum, with the thorough background knowhow. Along the way SLT should work more on identifying ways and means for improving its processes and related activities in order to be reached by the end customer in the most influential way. The company should plan and implement competitive, effective, reliable as well as practical awareness mechanisms according to the best representation of the influential factors identified, considering most to least force.

**Paper03: Factors influencing Broadband adoption in Sri Lanka : A study covering Western Province**

The findings from this research provide evidence for seven factors based on attitudinal beliefs, normative beliefs and control beliefs of households are influencing broadband adoption in western province.



MHz - Mega Hearts

Mbps Mega bitss per second

**Paper4: Predicting Consumer’s Complaint Behavior in Telecom Service: An Empirical Study of India, Sri Lanka, and Bangladesh**

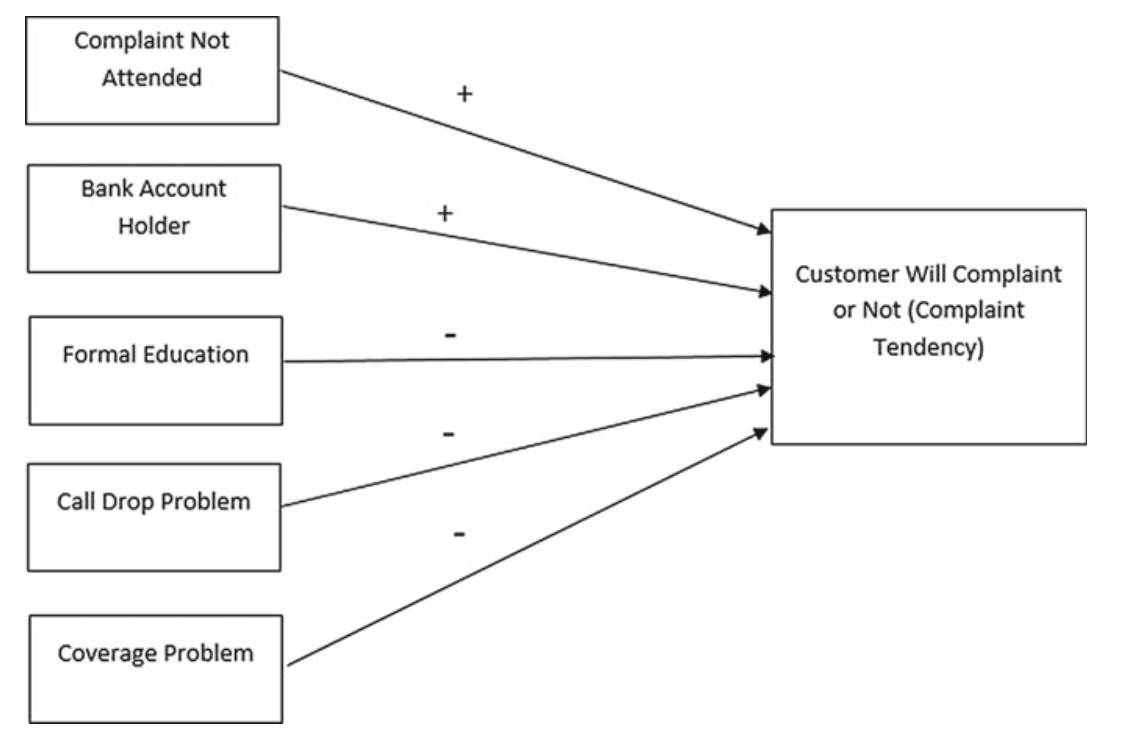
They aim to increase customer satisfaction to make their customer base loyal. Loyal customers stay longer with the company and are the good sources of revenue.

Whereas, on the other hand, a dissatisfied customer not only switches to other company but also spreads a negative “word of mouth” image hurting firm’s reputation.

management can know if their customers are satisfied or not from their **exit and voice.** Exit refers to when customers stop using the product/ service of the company and voice refers to the complaints of the customers.

Methodology: Predictive model using logistic regression

Binary class classification: Will customers complain or not?



Independent variables used:

• Number of SIMs—has significant correlation with the other variables. More

number of SIMs indicates that customer uses service more.

• Complaint not attended—is clear indicator of customer satisfaction. If telecom

operator do not attend complaint properly customer may switch to other telecom

operator. This variable has positive correlation with complaint tendency.

• Change in service provider—unsatisfied customer change their telecom opera-tor. This dissatisfaction comes when telecom operator do not handle complaint

properly. This variable has positive correlation with complaint tendency.

• Age—Age has negative correlation with the complaint tendency. It means older

people do less complaints compared to the younger people.

• Bank account holder—people have bank account and they more likely to

complaint more in compared to the customer who do not have bank account.

Also, it has positive correlation with the complaint tendency.

• Formal education—has negative correlation with the complaint tendency. It

shows that people having more education complaint less.

• Nature of mobile phone—smart mobile phone owners complaint more as this

variable has positive correlation with complaint tendency.

• Customer satisfaction—correlation is also positive.

• Number of recharge per month—has negative correlation with complaint

tendency.

• Direct Contact—negative correlation.

• Contact through call center—customer contacts with the telecom operator

through call center his tendency to complaint is more.

• Coverage problem and Call drop problem—negative correlation with the

complaint tendency

• Gender—has positive correlation.

• Private telecom company—has positive correlation.

Findings and Conclusion:

customer with bank account complaint is more compared to

nonholders. Also, those who have complained about banking services have also

complained about telecom services. Less educated customers are complaining

more. Unlike, educated customers they might take time to exit. Call drop and

coverage related issues are main problems that trigger complaining behavior.

**Paper05:A Descriptive Analysis on Digital Behaviour of Young Adults in Sri Lanka**

In Sri Lanka youth which consists of 15-19, 20-24 and 25-29 age groups have the highest digital literacy accounts for 76.6%, 77.9% and 71.1% respectively as at June 2019. Thus, it is evident that Sri Lanka is also following the global trends with regards to digital behaviour. [Guruge et. al., 2020] As the digital behaviour of youth in Sri Lanka is growing significantly, need for telecommunication services in a satisfactory level for the youth, has arisen.

Research approaches:

Quantitative

Methodology:

Conduct a survey

female and male young adults living in the Sri Lanka was taken as the population

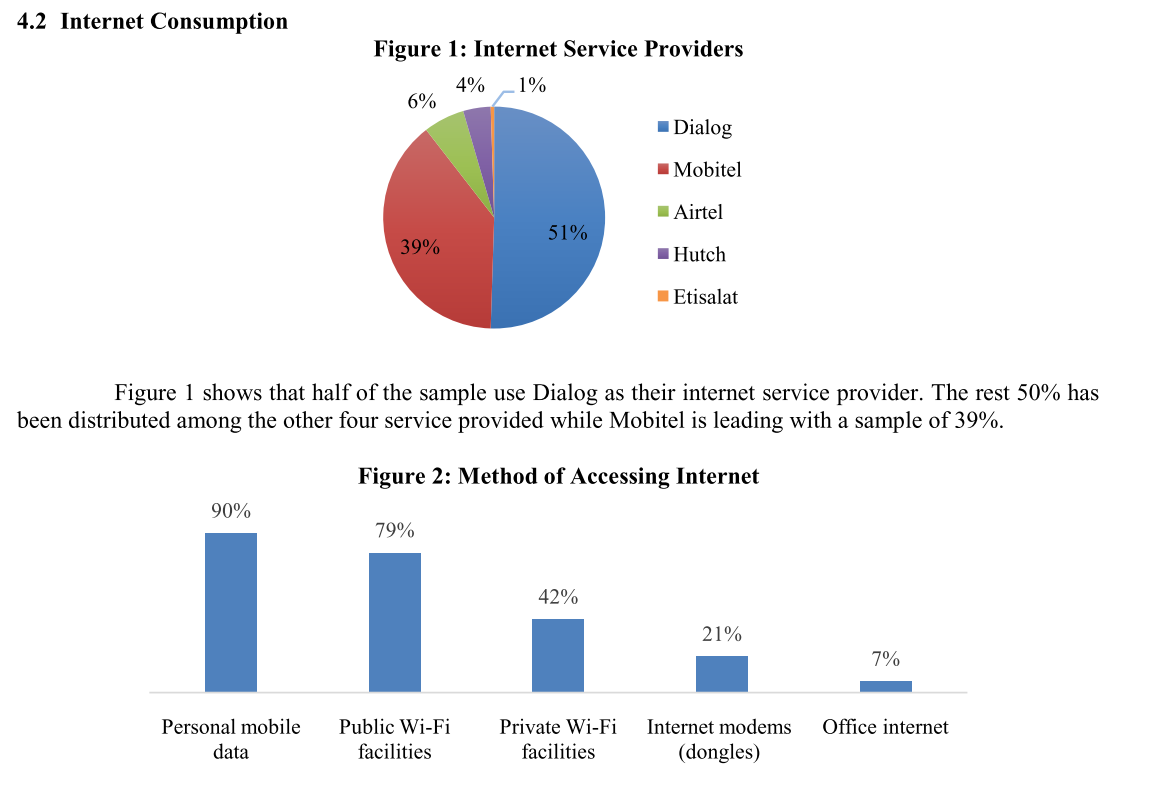
Sample:

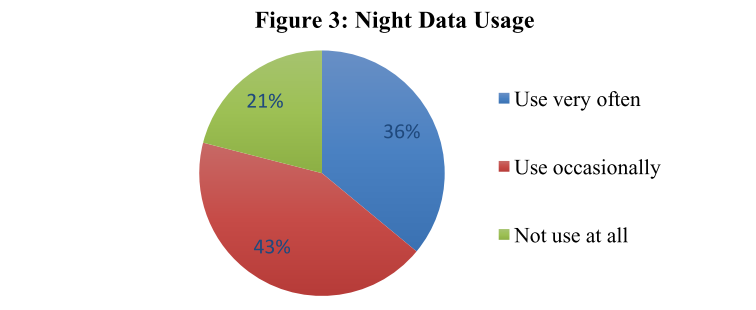
any young adult aged between 15-24 years, who is perusing either secondary or tertiary studies and living in any district in Sri Lanka is considered as the sample.

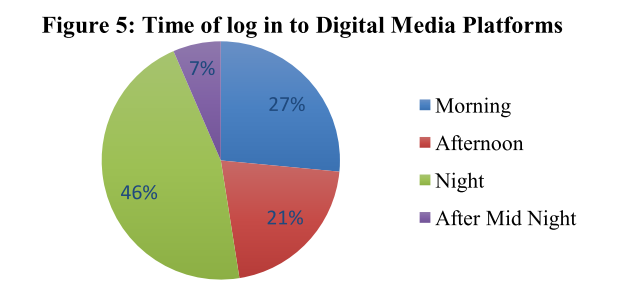
Findings:

research findings reveal that the majority of the young adults who pursue either secondary or tertiary education, log on to digital media platforms in a daily basis by their smart phones (91%) through consuming personal mobile data (90%) and they log in to online platforms while travelling. Furthermore, the frequently accessed digital media

platforms are Facebook (98%) and YouTube (93%) and mostly accessed in between 6-8 am and 6-10 pm.







out of the sample less than half of the respondents use night time data whereas

more than one third use night time data occasionally. Furthermore, one in every five does not consume mobile data at all.

**Paper06: Determinants of Customer Satisfaction in Telecom Industry - A Study of Indian Telecom Industry**

Methodology: study through a survey. Target group: university students

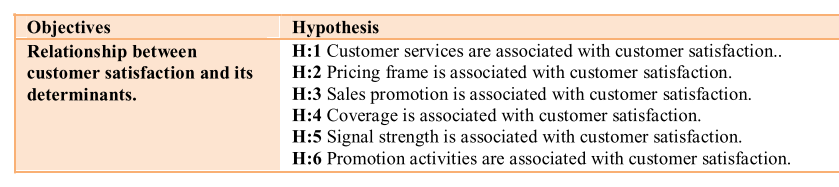
Satisfaction can be defined as a

features or characteristics that can full the either a need or want of a consumer in better way than competitors.

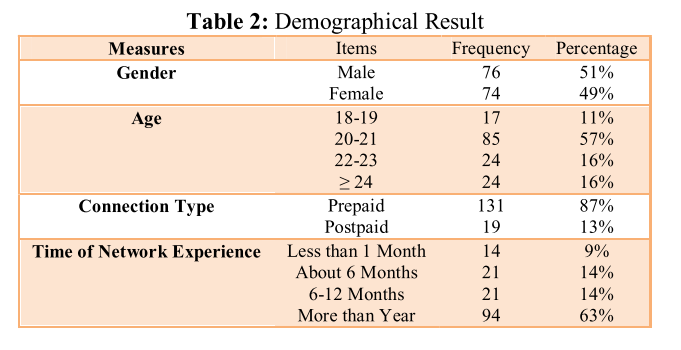
there are strong relations between satisfaction and loyalty

Factors affecting customer satisfaction: customer service, price fairness, sales promotions (ex: for voice services extra free minutes, limited offer to get double balance, free VAS, MMS at low rate etc), coverage of services,





Considered socio-demographic factors:



Findings:

It is concluded that price fairness and coverage are the key factors contributing towards customer satisfaction of University students. Keeping the findings of this research it is very obvious that the companies should always continue to emphasize on price fairness and coverage for consumer satisfaction in telecom industry. Coverage also influences customer satisfaction. The problem of coverage is generally in rural area where sometime customers are not able to gain services from any particular service provider. That factor can compel the customers towards brand switching in order to get the coverage in any particular area. Furthermore, customer services also impact the customer satisfaction regarding any service provider.

**Papaer7: PREDICTING POPULATION-LEVEL SOCIO- ECONOMIC CHARACTERISTICS USING CALL DETAIL RECORDS (CDRS) IN SRI LANKA**

Call Detail Records (CDRs) can broadly describe three dimensions of human behavior: social networks, consumption activity, and mobility

leveraged CDRs and airtime credit purchase data to infer the relative income of individuals, based on the

assumption that those who made larger airtime credit purchases were relatively more affluent that

those mobile users who made multiple purchase of smaller airtime credit.

CDR data captures the following: (1) A unique identifier for the calling/sending party; (2) A unique

identifier for the other party on the call; (3) The date and time at which the event was initiated; (4) The

ID of the cellular antenna the subscriber was connected to at the time of the call.