KONOWLEDGE DISCOVERY AND MANAGEMENT

PROJECT PROPOSAL

Team members:

VISHNU GUDE

VEMPERALA Y V AKHILESH

AYYAPPA KUMAR NEERUKONDA

LAXMAN DUTT DEGALA

In this course we are going to develop a mobile application and web application based on Hadoop technology which can access the datasets available to us.

**Motivation:**

In today's world efficient way of power consumption has become an big issue. Now a days it is every one's responsibility to save electricity because electricity is non-renewable energy. Even though many of us have the intension to save this energy, but without having the much knowledge about how much energy is consuming on every day, and by every appliance, we are unable to do this. By knowing the basic knowledge like consumption made by many appliances like refrigerator, micro wave Owen, fan, television and computer or laptop etc.....and by comparing it with the average consumption by these devices on a regular basis one can actually control the usage of power. So our main motivation is creating an mobile application which provides the information to the user regarding the electricity consumption. on every day and on every appliance and compare it with average consumption of that particular appliance.

**Significance:**

The purpose of this project is to create a mobile app that helps to educate the people or users in a particular area about the power consumption made by different appliances in their respective household and by comparing it with the average consumption of that appliances in a graphical manner that help the user for the efficient usage of electricity and also by storing the data user can actually compare it with previous consumption made by that appliance and to know how much he saved or wasted on that appliance. In this project we have data of one million users in a society in excel sheet. This data is stored in a database and accessed by using Hadoop and web application.

**Objectives:**

\*With this application we are going to provide the user awareness regarding how much energy is consuming every day by every appliance.

\*Providing graphical representation of user power consumption on daily wise, monthly wise, average daily consumption average monthly consumption.

\*utilization of smart meter provides the information regarding power consumed by every appliance connected to it.

\* energy provider can also provide suggestions to the user, So that user can regulate his power consumption

\*By knowing the power consumed by every appliance he can better regulates the power consumption of every appliance.

**Features:**

1.Graphical view of energy consumption.

2. Will try to represent things mostly in diagrammatic manner to have an idea about it.

3.Special feature is web application that update the data base.

4. Application wise information and their energy consumption.

5.To use an hadoop technology that can access database having one million users.

**Requirement Specifications:**

Objectives for discovery of patterns and adaptive self-learning modeling on energy consumption data that can accomplish the following:

1. Identify potential features affecting energy consumption in US population since 1997 up to 2009.
2. Recognize minimal, optimal and wasteful energy use patterns.
3. Search patterns by clustering household observations into corresponding prominent groups based on associated energy cost.
4. Discover evolving features, which influence energy consumption metrics as a function of time.
5. Predict energy expenditure (dollars) and energy consumption (BTU).
6. Determine energy use based on individual choices for saving energy and saving cost.
7. Present interactive time series graphs for visualizing energy consumption from various appliances and mains supply of the household.

Discovery and analysis of feature vectors for recognizing evolving and static parameters was the main concern, and this identification problem has to be addressed.

The modeling strategies used for the aforementioned goals are:

1) Exploratory analysis of data for discovering evolving features,

2) Principal component analysis for dimensionality reduction on single node analysis,

3) Context aware adaptive clustering to extract patterns, and

4) Prediction of energy consumption over minimum cluster, optimal cluster and maximum energy cluster patterns.

A learning methodology for predicting future energy demands, which involve modifying substantial features, discovering new patterns for efficient energy usage and recommending users per their needs, is proposed as a solution. Building on top of a distributed file system, the energy saving recommendation system is self-evolving while compiling efficient energy usage patterns and relevant feature information from historical data and current consumption metrics.

**System Architecture Diagram:**

Architecture diagram of the system

**Domain Sources:**

Different Models for Energy Estimates

Several statistical models based on user’s present energy consumption have been developed, some of which closely monitor a household’s daily activities for energy management. Most of these works are centric to very small population and provide instant results only on basis of recent data.

1. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a nonparametric method for finding the relative efficiency of Decision Making Units (DMUs), each of which is an entity responsible for converting multiple inputs into multiple outputs. DEA has been applied to a variety of managerial and economic problem situations in both public and private sectors. DEA assumes each DMU uses multiple inputs to yield multiple outputs and defines the process, which changes multiple inputs into multiple outputs as efficiency score.

2 Hidden Markov Model

Hidden Markov Models (HMMs) are commonly used in many fields, e.g., speech recognition, pattern recognition and gesture recognition. HMMs are stochastic models for energy data that are serial or temporal. The word "hidden" in HMM refers to the hidden states that are mapped to the data. This model is typically used for modeling sequences of events. This model is particularly useful when the data is noisy and incomplete. It is based on estimating probability distributions and efficient algorithms for learning and recognition such as Baum-Welch and Viterbi Path Counting algorithms.

3 Time Series Analysis

Xianghui analyzed time series systems and set up a time series artificial neural network model to predict energy output. Predictions were based upon the level of payment for accounting periods based on the Subotica Electric Power Distribution Enterprise’s data, where the consumer’s time series data analysis in the category of households was derived using their consumption and payments for some accounting periods from 2001 to 2006. Results were drawn from analyzing electric energy at higher rate or lower rate consumption, obligation, payment and average temperature in the region.

**Methodologies and Algorithms:**

Pattern and Rules Extraction Algorithm

Extracting the adaptive patterns and rules that reason the energy consumption level of a household is critical to the logistics of energy management’s adaptive technique. The patterns are used to intelligently monitor and control energy usage in a household. In order to extract rules and arrange features of a household in order of their priorities for saving energy, we designed a fuzzy clustering based on evolving features to extract the underlying structure of household energy consumption data. For taking into account the effects of evolving and correlated features, a reverse adaptive loop is added to clustering process. Hence, the consolidated name of clustering methodology is Adaptive Reverse Fuzzy Clustering Algorithm.

After clustering, rules are inferred and patterns are extracted from all households that belong to the same cluster using association rule mining. The mining process is used to find rules with defined precision and confidence values.

**Adaptive Reverse Fuzzy Clustering:**

Adaptive Reverse Fuzzy Clustering is an adaptive methodology for fuzzy partitioning of households based on feedback of previous and present energy consumption data. The method classifies energy consumption behavior of a household into three categories: minimum energy using households, maximum energy using households and medium energy using households.

In energy data, the household population show diverse energy consumption behavior, which ranges from optimum energy usage to excessive or wasteful energy usage. We justified wastage in energy use pattern of a household, by comparing every two households having equivalent static features like same number of members, same geographical location, same household structures and dwelling conditions in comparable economic zones, but with wide differences in energy bills and energy use behavior. Such households are grouped based on similar energy spending and usage patterns into same clusters. The one having maximum energy spending is placed under maximal energy cluster zone and the one having a smaller energy spending is placed in an optimal or minimal zone, based on threshold values for each cluster and the average number of members in each cluster.

**Technical Requirements:**

The user interface development for mobile devices like smart phones, tablets and touch screen monitors makes use of Android development environment, where the module for user interaction and decision making is programmed using Model-View-Controller design pattern. For Android development, programming languages Java and JavaScript are used for intelligence control, HTML5, Phone Gap and CSS3 are used for styling and designing the layout of user interface.

All decision making (data mining and streaming) services are deployed in cloud. Java is used as the programming language for designing the interface layer of android that interacts with visualization web services. R and Mahout are used for statistical modeling. Statistical modeling for designing the data model and the recommendation system also makes use of R machine learning tool for exposing the priorities of features to users via browser interfaces.

Following is a list of tools and technologies used for creating the system

1. Programming Languages: Java and R
2. Statistical Modeling software and analytical workbench: R Studio and Octave
3. Mobile User Interface device: Any Android 4.0+ based system with touch screen capability or smart phones
4. Operating System: Android 4.1+ for android devices, windows for programming Android applications in Eclipse studio, Linux instances in cloud for deploying web services and Hadoop distributed storage system.
5. Wireless Support: Wi-Fi / 4G for internet connectivity
6. Cloud Resources: Cloudera Hadoop Instance and windows server for hosting web services
7. WC3 Standards: XML, XML Schema, RDF, RDF Schema for defining interfacing layer
8. Ontology Specifics: OWL (Protégé), Jena
9. Distributed Computing: Hadoop for distributed storage and Mahout for predictive analytics. We use Apache Flume, Hadoop HDFS, Apache Oozie, and Apache Hive for collecting streams of mains supply, environmental and climatic changes.

**Development Tools:**

1. Eclipse IDE
2. R studio and R software
3. Android Software Development Kit, and emulators for developing Android application

**Existing Applications:**

Existing Energy Management Applications

Energy and resource conservation in home settings have been a consistent area of social and environmental research since 1970 [4]. Energy management is a broad concept and problem domain. Please find below two main approaches are proposed to provide energy management.

1 Energy Disaggregation

Energy disaggregation is a non-intrusive load monitoring technique, which involves using a single energy meter to measure the energy consumption of an entire home. It then uses machine learning and signal processing techniques to reverse engineer appliance usage profiles in the house. Most of the research in this area made use of supervised machine learning to infer appliance usage from the aggregate energy consumption. We used collected energy consumption from mains supply of a household to predict the household’s energy in long run (i.e., weekly, monthly, and yearly). The challenge in energy disaggregation is that individual appliances have very different energy signatures that are hard to spot unless high-resolution meters are available. Moreover, even if some appliances can be accurately recognized and classified given an existing training set using machine learning techniques, such solutions perform poorly when subjected to new instances under the same category

2 Demand Response System

According to the Federal Energy Regulatory Commission, Demand Response (DR) is defined as “Changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized.”

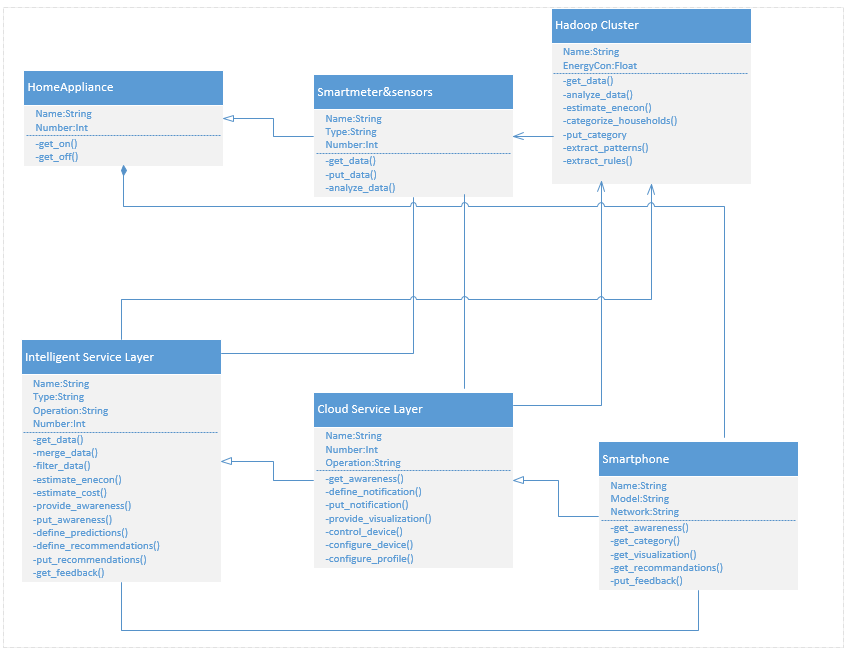
The basic limitation of DR lies within its targeted end user, which is only the building manager. A lack of control scheme for users other than managers or enterprise building managers makes DR unsuitable for residential usage. There is no automated intelligence or recommendation for users to switch plug-loads from devices, which are off to devices, which demand energy.

In electricity grids, demand response (DR) is similar to dynamic demand mechanisms to manage customer consumption of electricity in response to supply conditions; for example, having electricity customers reduce their consumption at critical times or in response to market prices. The difference is that demand response mechanisms respond to explicit requests to shut off, whereas dynamic demand devices passively shut off when stress in the grid is sensed. Demand response can involve actually curtailing power used or by starting on-site generation which may or may not be connected in parallel with the grid.

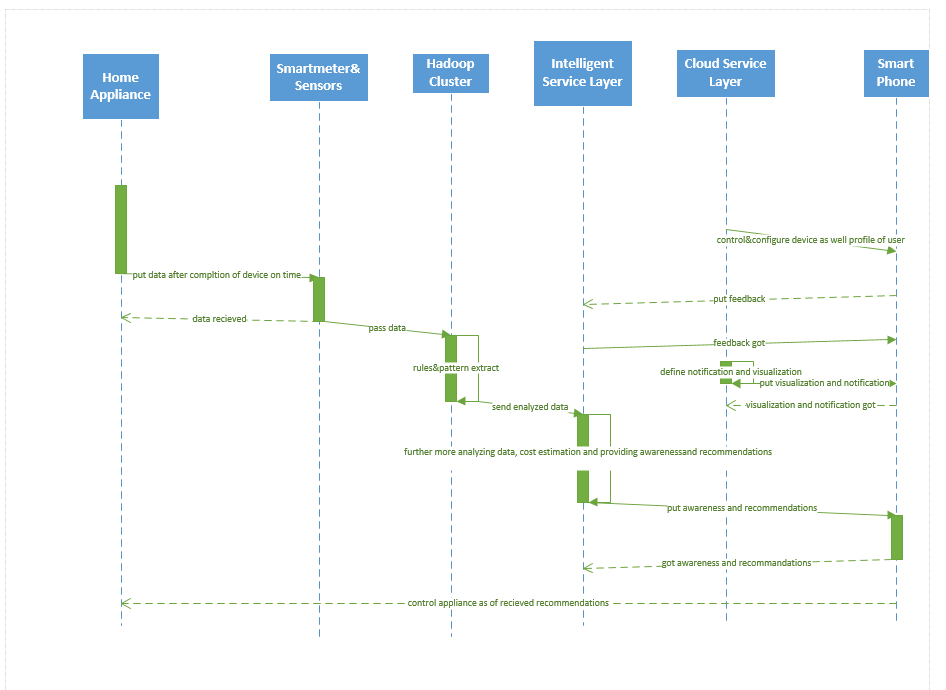
**New Application to be built: CAAM – Context Aware Adaptive Model for Smart Energy**

Both Demand response and dynamic demand devices are limited in their scope of automation and usage for end users. CAAM (Context Aware Adaptive Model) provides both automated mechanisms for energy savings and interactive visualization of energy use for real time decision making.

Context aware adaptive fuzzy clustering is an adaptive model for predicting future energy demands and preparing present households to adapt to means and techniques to better utilize electrical energy. An approach to discover patterns for saving energy, followed by a design of a predictive model that can estimate energy consumption and provide adaptable recommendations is described in the following sections. The evolving model learns and adapts to changing features and data values. It is based on past energy consumption metrics and the present scenario of energy usage.



UML CLASS Diagram for proposed system



UML SEQUENCE Diagram for proposed system

Model is context aware and is based on contextual information such as:

• Demographics of a household, e.g., location and housing unit or apartment

• Occupying member’s information, e.g., age, purchasing balance, and preferred temperatures,

• The structural features of the house, e.g., the number of bedrooms, doors, windows, floors, type, built year, material used for building and potential environmental properties related to a household,

• Accommodating changes in place and location

• Location-specific weather and climatic changes

**Reference:**

Based on Thesis done by SwatiSoni