Aufgabe

* Interpolation von Werten
* Prognosen für weitere Werte
* Schätzen von weiteren Daten anhand Profils
* Daten = 2 Lastgänge im Excel Format (2016 & 2017):

|  |  |  |
| --- | --- | --- |
| Datum | Uhrzeit | Wert [kWh] |
| 01.01.16 | 00:15 | 0,228 |
| 01.01.16 | 00:30 | 0,211375 |
| 01.01.16 | 00:45 | 0,192375 |

Vorbereitung

1. Daten aus Excel Dokument extrahieren
2. Daten in geeignete Datentypen konvertieren
3. Daten für Visualisierung vorbereiten:
   * Hilfsmatrizen
     + Stündlich
     + Täglich
     + Monatlich
   * Anpassungen

Algorithmen

2.1 Korellationsmatrix

Eingabe: mList = durchschnittliche Monatshilfsmatrix  
 Ausgabe: aList = Korellationsmatrix  
  
 mList = outlier(mList) // geglättet

FÜR JEDES jahr:

FÜR i = 1 BIS 12

FÜR j = 1 BIS 12

aList[jahr][i][j] = mList[jahr][i].value / mList[jahr][j].value

* Reduce operation -> averaging

Text:

Einleitung:

Wir drei hatten generell die Aufgabe, Werte zu interpolieren. Es gab eine Masterarbeit, die hat sich mit der Rentabilität einer Solaranlage beschäftigt. Früher hat sich eine Solaranlage sehr schnell rentiert, da man sämtlichen Strom verkaufen konnte. Dabei war der Erlös höher als die Kosten, die man für die Nutzung von Strom hat. Mittlerweile ist es allerdings so, dass der Erlös niedriger ist als die Kosten, daher sollte man gucken, wie viel Solarstromgenerierung zu Zeiten möglich ist und wie hoch der Eigenverbrauch ist, damit man diesen benutzt, statt Strom von Fremdanbietern zu beziehen. (Da nicht immer Sonne vorhanden ist, bezieht die Masterarbeit ebenfalls in Betracht, einen Akku zu besorgen). Für die Berechnungen allerdings braucht die Masterarbeit Stromverbräuche, um möglichst genau zu ermitteln, nach wie viel Zeit sich die Investition rentiert.

Aufgabe:

Unsere Aufgabe besteht es also darin, dass wir möglichst realitätsnahe Daten generieren. Im besten Fall kann ein Interessent einen kompletten Lastgang zur Verfügung stellen und die Berechnung wird ziemlich akkurat. Allerdings ist dies nicht oft der Fall. Wenn Daten vorhanden sind, dann sind die oft nicht komplett – entweder sind nur Teile vorhanden (z.B. August 18 – Januar 2019), oder es kann auch sein, dass das Strommessgerät mal ausgefallen sind und plötzlich irgendwelche Lücken vorhanden sind, wenn der Ausfall vom Gerät nicht sofort bemerkt wird. Daher besteht unsere erste Aufgabe darin, die Werte zu interpolieren. Eine Solaranlage rentiert sich natürlich nicht sofort innerhalb von einem Jahr, es braucht mehrere Jahre. Der Stromverbrauch kann sich allerdings ändern, daher müssten ggf. Prognosen für Folgejahre erstellt werden, oder, wie im Beispiel vorhin: Januar -> Mai 2018, da fehlen die Monate bis August 19. Zuletzt gibt es noch den Fall, dass die Verbraucher gar keine Lastgänge haben, die sollen dann die Möglichkeit bekommen, Verhalten und Verbräuche zu schätzen anhand eines Profilgenerators.

Für das Projekt liegen uns zwei Lastgänge, eins für das Jahr 2016 und eins für das Jahr 2017 vor. Was für ein Haushalt das ist (Details / (Anzahl an Personen, Arbeitsstellen, Alter)) liegt uns nicht vor. Sie liegen in Form einer Exceltabelle vor, bei denen jeweils ein Datum, eine Uhrzeit und der gemessene Wert in kWh vorhanden ist.

Vorbereitung:

Zunächst extrahieren wir die Daten aus der Exceltabelle. Dabei lesen wir die Datei ein und legen die Werte entsprechend in passende Variable Typen ab (z.B: Datum in ein Date Objekt, generell die gesamten Daten in einer (Tree-)Liste). Ich / Wir finden es oft hilfreich, die Daten auch anschaulich zu betrachten und haben dafür verschiedene Visualisierungen erstellt. Dafür haben wir Hilftsmatrizen generiert, die jeweils durchschnittliche Werte der Stunden, Tage & Monate haben. Dabei haben wir alle Daten durchlaufen und jene, die beispielsweise in derselben Stunde waren (also 00:00, 00:015, 00:30, 00:45) miteinander addiert und durch die Anzahl geteilt, selbiges für Tage & Monate. Für unsere gewählte Visualisierung mussten die Werte noch in ein passendes Format angepasst werden.

Korellationsmatrix:

Der erste Gedanke war herauszufinden, wie die Monate abhängig voneinander sind – pauschal wird von verschiedenen Seiten eine Spanne von 10 bis 15% Unterschied von den Sommer- & Wintermonaten genannt. Daher haben wir also erneut die Monatshilfmatrix berechnet, die abhängig von der Tagshilfsmatrix ist. Bei der Tagshilftsmatrix allerdings der Unterschied, dass wir Werte, die aus der „Reihe tanzen“, abgeschnitten haben. Zunächst haben wir einfach alle Werte eines Tages genommen, sortiert und die ersten 10 oberen und unteren abgeschnitten, um die Daten zu glätten. Später allerdings haben wir stattdessen den Boxplot Algorithmus genommen, um lediglich nur auf die starken Ausreißer zu wirken. Danach alle Tage eines Monats, wie bereits vorhin kurz erklärt, nur noch aufaddiert und durch Anzahl der Tage im Monat geteilt.

Jetzt: Da wir zwei Lastgänge haben, haben wir für jedes Jahr eine 12x12 Matrix aufgebaut. Wir durschlaufen also alle Jahre (hier 2), durchlaufen dann für i von 0 bis 11, darin für j von 0 bis 11 und teilen dann die jeweiligen monatlichen Durchschnittswerte mit den anderen Monatsdurchschnittswerten. Am Ende führen wir noch eine Reduce Operation aus, d.h. wir wollen alle (hier leider nur 2 vorhanden) Monatshilfsmatrizen aufaddieren und durch die Anzahl teilen, um die Durchschnitte zu bekommen. Zwar haben wir Ausreißer bereits eliminiert, die Lastgänge an sich können allerdings auch Ausreißer sein (wie z.B. ein Haushalt, der während der Winterzeit in Afrika wohnt). Je mehr Lastgänge man hier also inkludiert, desto genauer werden die Relationen.

/\* Dann zeige ich kurz die Seite \*/  
Hier kann man z.B. den Graphen sehen, der zeigt gerade die monatlich gemessenen Durchschnittswerte an. Wir können zum Testen mal den Februar entfernen und durch die Korellationsmatrix versuchen, die Werte zu interpolieren (dabei berechnet er die Werte von den vorhandenen Monaten \* den Korelationsmatrixwerten / anzahl von vorhandenen Monaten). Ergebnis hier ist zwar ziemlich genau, aber bringt nicht so viel, da es im Moment sowieso sehr von den Lastgängen abhängt (da nur zwei „Trainingsdaten“ vorhanden waren).   
Hier kann man die Korelationsmatrix sehen. Was einem auffällt ist z.B. dass der Dezember im Vergleich zum August, zumindest bei den Daten, die wir haben, 190% sind, also 90% mehr.  
Dann gibt es noch einen kleinen Profilgenerator, einfach – kennt jeder von den ganzen Stromseiten. Spannend ist eher das Erweiterte. Hier kann man Geräte auswählen wieviel die Verbrauchen, wie viele davon und wie lange die an sind . Dann bekommt man einen jährlichen Verbrauch, und dank der Korelationsmatrix kann man den auf die verschiedenen Monate aufteilen. Auch kann man die Daten visualisieren, dabei wurde hier geguckt, wie viel ein Tag bei den Testdaten vom Monat ausmacht und dementsprechend bekommt der Tag diesen Anteil von den geschätzen Monatswerten.