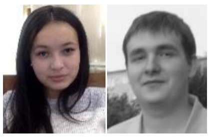
Methods of aggregation different data for estimating and evaluating potential of renewable energy

A Shakhmetova, K Yakunin

Faculty of Information Technology, Kazakh-British Technical University, Tole Bi Str. 59, 00500 Almaty, Kazakhstan

**Corresponding author's e-mail: asima-1995@mail.ru*



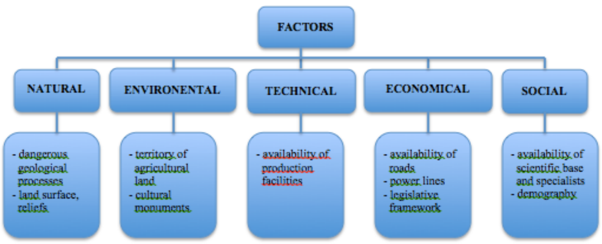
Abstract

It is important to be able to evaluate the territory, in order to establish some kind of generator. There are many factors influencing on installation, therefore article below illustrates different ways of evaluating and finding optimal place. Each system of assessment has own advantages and disadvantages, all of them will be discussed and represented in examples. Finally, the most efficient system will be chosen for future usage in project.

Keywords: Bayes system of inference, apriori probability, aposteriori probability, conditional probability

1 Introduction

The usage of renewable energy source allows improving ecological condition of territory and improving the stability of energy supply. The main problem of establishing the generator is to find optimal location, which ensures the maximum amount of energy received at minimum cost as well as the economical feasibility, which includes analysis of infrastructure, local economic level and proximity of potential energy consumers. Engineering considerations are not enough to determine the optimal place, because wild range of factors affect placement of such structures. Therefore, data will be aggregated to obtain an integrated assessment. It is proposed to use Bayesian inference system or Multiplicative Multifactor Model, which is able to get the evaluation without accurate, detailed data (lack of details). Bayesian system gives two answers simultaneously, namely "for" and "against" with different probabilities. Also, the ratio between results is considered to be "degree of certainty" of the algorithm – higher ratio means more "certainty" and generally means more detailed (and less ambiguous) data was available.



2 Multiplicative multifactor model

For aggregation of data could be used method based on summation of factors or a multiplicative method. The multiplicative multifactorial model is represented by the formula:

$$F = \prod_{i=1}^{N} f_i$$

This model aggregates all the data as a product of the values of factors. If the specific factor does not affect on assessment, its value is set to 1, if it has a significant negative impact, to 0. There is a problem of reliability of results when multiplicative multifactor model is used. The reason of that is that some data could be unreliable and not accurate.

3 The Bayes system of inference

The multiplicative evaluation system is not accurate, in a way it could not provide the opportunity to assess the quality of obtained data, especially in case of lack of data or incomplete confidence in their reliability. In this case, the following stochastic rule is used:

"If h is true, then e will be observed with some probability p". For instance, h could be considered as an event, which indicates the installation of a generator, and e is an event which determines an evidence of correctness of decision. In other words, h is denoted as an event, which means that hypothesis is correct, when e is an event which could confirm the correctness of hypothesis.

The possibility of h, in case of e occurs could be expressed by the following formula:

$$p(h|e) = \frac{p(e|h) \times p(h)}{p(e|h) \times p(h) + p(e|\backsim h) \times p(\backsim h)}$$

$p(e|h)$ - conditional probability of the occurrence of the event e, when h is valid

p(h) – prior probability of the hypothesis h
p(e│∽h) - conditional probability of e, when h is not valid
p(∽h) – probability, that the event h is not true, according to the formula of total probability, it can be calculated by the formula:

$$p(\backsim h) = 1 - p(h)$$

In this way, in order to calculate the conditional probability p (h│e), it is nessecery to know the probabilities p(e│h) and p(e│∽h), also apriori probability p(h). It is possible to consider several events {h1, h2, ..., hn} ⊂H.

4 The algorithm consists of following steps:

For each hypothesis h⊂H and all evidences e⊂E, it is necessary to determine a priori probabilities p (h) and conditional probabilities p (e│h) and p (e│∽h), followed by calculation of possibilities p (h│e) and p (h│∽e) for each hypothesis h and evidences e. For each subsequent evidence, the prior probability of the hypothesis p (h) is set equal to the conditional probability p (h │ e) found at the previous step. As a result, two estimates p (h │ e) and p (h │ ~e) will be obtained for each hypothesis, which can be compared with each other and chosen those p (h │ e) and p (h │ ~e), which will have the significant difference between maximum and minimum values.

For example, there is plan to install a large wind generator in certain place. However, a large number of factors have influence on the decision, such as roads, railway stations, favorable engineering-geological conditions of the area, arable land, hayfields, rural or urban areas, forests, pastures and pastures, sands, gardens etc. Each of the factor has the conditional probability value 0<=P<=1. The bigger value is, the more suitable the area is for establishing a generator. You can find two examples of calculations using the model below. The values of p(h1) and p(h2) start of at 0.1.

TABLE 1 Example of calculation using the model

p(h1)	p(h2)	p(e|h1)	p(e|h2)	p(~e|h1)	p(~e|h2)
0.1	0.1	0.9	0.5	0.1	0.5
0.166667	0.05814	0.5	0.5	0.5	0.5
0.166667	0.05814	0.9	0.5	0.1	0.5
0.264706	0.033156	0.9	0.5	0.1	0.5
0.393204	0.018696	0.9	0.5	0.1	0.5
0.538405	0.010474				

The tables above illustrate several properties of the model. Value of 0.5 is considered to be neutral value and doesn’t affect p(h1) or p(h2). It might be counter-intuitive from probability theory point of view, however it allows to estimate positive and negative effect of different factors in a simple and intuitive way.

TABLE 2 Example of calculation using the model

p(h1)	p(h2)	p(e|h1)	p(e|h2)	p(~e|h1)	p(~e|h2)
0.1	0.1	0.9	0.5	0.1	0.5
0.166667	0.05814	0.5	0.5	0.5	0.5
0.166667	0.05814	0	0.5	1	0.5
0	1	0.9	0.5	0.1	0.5
0	1	0.9	0.5	0.1	0.5
0	1				

The value of 0 works as a hard boundary – if one of the factors with p(e|h1) = 0 is present, establishing of generator is completely impossible and other factors cannot change it.

4 Conclusion

Methods of aggregation different data for estimating and evaluating potential of renewable energy were researched, a number of factors were analyzed and systemized and two models of evaluation of feasibility of installing a generator were considered: namely, multiplicative/additive models and Bayesian system of inference. From the initial research it became clear, that multiplicative MMM has a serious flaw – it is almost impossible to tune to work with heterogeneous and unreliable data. On the other hand, Bayesian inference can deal with this problem due to its stochastic nature, which in addition represents the probabilistic nature of the object of the research. From initial research, it can be deduced that Bayesian inference can be used to estimate feasibility of establishing generators.

Acknowledgments

The work was funded by grant No. 0168/GF4 of the Ministry of Education and Science of the Republic of Kazakhstan.

References

[1] Muhamedyev R I, Mustakaev R, etc. 2016 Aggregation of dissimilar data in the decision support system for the use of renewable energy sources *Materials of the International Scientific Conference "Informatics and Applied Mathematics" Almaty* 322-30
[2] Bernardo J M, Smith A F M 1994 *Bayesian Theory* Wiley
[3] Lee P M 2012 *Bayesian Statistics: An Introduction* Fourth Edition John Wiley
[4] Samaniego F J 2010 *A Comparison of the Bayesian and Frequentist Approaches to Estimation* Springer, New York
[5] Bayesian Networks and BayesiaLab *A practical introduction for researchers by Stefan Conrady and Lionel Jouffe*