**TOPIC MODELLING OF REVIEWS GIVEN ON GLASSDOOR MANUFACTURED BY HONEYWELL**

***UCHECHUKWU FREDRICK OKONKWO***

Faculty of Engineering, Environment and Computing

MSc Data Science and Computational Intelligence

Coventry University, Coventry, United Kingdom

STID: 11738697

[*okonkwou2@uni.coventry.ac.uk*](mailto:okonkwou2@uni.coventry.ac.uk)

***Abstract –* This paper discusses the unsupervised topic modelling tasks carried out on the Honeywell glass door manufacturing company dataset which are positive and negative reviews given by the staffs and anonymous customers obtained from the Kaggle data repository. I would be employing two different topic modelling algorithms namely: Latent Dirichlet Allocation (LDA) and the Latent Semantic Analysis (LSA) to extract the abstract topics from the given corpus. The coherence score will be used to compare both methods to determine the best number of topics for each model.**

***Keywords*: Topic modelling, Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), positive reviews (pros), staffs, customers.**

1. **INTRODUCTION**

Employees and customers will unavoidably air out their experiences with the company, product or services rendered in form of reviews. Companies spend a lot of time and money to keep track of the comments given to protect themselves as well as to optimize the treatments to the staffs and improve generally the quality of their service to potential customers. Topic modelling is a powerful tool used in natural language processing (NLP). It helps the company to analyze a large textual document to find out what the document is talking about. In NLP, the latent Dirichlet allocation method is a generative statistical model that allows sets of observation to be explained by unobserved groups that shed light on similarities of the textual data by discovering topics and then automatically classifies them within the collection in terms of how relevant these individual documents are to each of the discovered topics. LDA assumes that the words found in each document are related. LDA consists of two parts, the words within a document (know factor) and the probability of words belonging to a topic which is required to be calculated. The algorithm tries to determine how many words belong to a specific topic for a given document. Additionally, it attempts to determine how many documents belong to a specific topic due to the existence of a certain word. LSA, also known as latent semantic index (LSI), on the second hand is also a topic modelling method used for concept searching and automated document categorization. It has several drawbacks and the major one of them is its inability to capture

multiple meaning of words. LSA learns latent topics by performing a matrix decomposition on the document-term matrix using singular value composition which could be understood as typical dimension reduction or noise reducing technique. I would be using both the LDA and LSA methods then I would compare the topic coherence measure which is a widely used performance metric to evaluate the topics models.

Reviews make a great distinction in how consumers/customers/users feel about a product of a company. It could tell how much they are satisfied or not after the purchase and use of a particular product which could help the manufacturing companies identify ways to improve the quality of their products/services. This also applies to the reviews, both positive and negative comments, of the employees in a company.

The problem is how do they understand how the customers or employees really feel based on these reviews recorded as “pros” and “cons”.

1. **DATASET OVERVIEW AND PRE-PROCESSING**

***2.1 Data description***

I would be focusing on the reviews (“pros”) given by the employees who are working or have worked with Honeywell manufacturing company. The dataset helps us understand the important topics discussed by the staff. This dataset contains 2000 rows with five columns with the headers: serial number, date review was given, title (job title of the employees), pros (majorly positive comments) and finally the cons (majorly negative comments). The pros and cons respectively might contain a mixture of both negative and positive comments from these employees. This data was gotten from the Kaggle data repository website.

A picture containing calendar

Description automatically generated

Figure shows a partial overview of the Honeywell staff review corpus

***2.2 Exploratory analysis***

Prior to the exploratory analysis, some pre-processing was done which is a very important step in language processing before parsing the textual document into the algorithm to achieve a valid result. Firstly, I imported the necessary libraries needed for the tasks. They are all embedded in the natural language tool kit (NLTK) library to remove stop words which are majorly 1-3 letter words , numbers, punctuation marks, tokenize as well as to lemmatization which is a process of grouping together the different forms of a word so they can be analyzed as a single item. Other libraries imported for the general task include: pandas for data frame operations, numpy, seaborn, genism which is an open source library used for NLP, pyLDavis, matplotlib for visualizations and the string.

A word cloud was plotted showing the most frequently used words appearing in the pre-processed corpus to help us understand the reviews by the employees concerning their work experiences.

Text

Description automatically generated

Figure showing the word cloud of most frequent appearing words

In figure 2 above the words in large fonts indicates they appear more. We see the words “great”, “work”, “good” appear very often as expected.

I also plotted the length of reviews against the number of words in each sentence. It helps us understand the distribution of this words and this is seen in figure 3 below.

A picture containing text, crossword puzzle

Description automatically generated

Figure shows the distribution of words in the reviews(pros)

Chart, histogram

Description automatically generated

Figure shows the distribution of words in the reviews (cons)

***2.3 Feature selection, filtering, and Vectorization***

Feature selection was inculcated so that the important words will be inclusive. This is done to reduce the number of words in the corpus that will be used for training to help its efficiency in time consumption in the training process. I set it such that we would have only words that are present in at least 3 different documents and in less than 50% of the documents are tagged necessary.

The unique number of words from the pros (majorly positive review) after pre-processing, this is before the feature selection (filtering) amounts to 2285. The unique number of words after the filtering amounts to 894 respectively. For the cons (majorly negative reviews), unique number of words after pre-processing , before filtering is 4192 and after filtering it amounted to 1648. Since these words can’t be fed directly into the algorithm, I used the term frequency and inverse document frequency to derive the statistical measure that evaluates how relevant a word is to a document by multiplying how many times a word appears in a document by the inverse document frequency of the word across a set of documents represented in numbers.

1. **METHODOLOGY**

As mentioned earlier, I will be applying two topic modelling methods to carry out this analysis.

***3.1 Latent Dirichlet Allocation (LDA)***

One most relevant application of LDA in machine learning is in topic discovering which is a sub problem in natural language processing. It discovers topics in a collection of documents then automatically classifies any of the individual document within the collection in terms of how relevant it is to each discovered topic.

LDA is a generalization of older approach of probabilistic latent semantic analysis (pLSA).

We have two metrics:

1. **θtd = P(t|d)** which is the probability distribution of topics in the documents
2. **Фwt = P(w|t**) which is the probability distribution of words in the topics.

So, we say the probability of a word given document, that is, P(w|d) is given by :

Where T is the total number of topics

For every topic tϵT, a multinomial distribution Фt is selected from a Dirichlet distribution with hyper-parameter β. For every document dϵM a multinomial distribution θd is selected from Dirichlet distribution with hyper-parameter α. For every word w discovered in document d, a topic zn is selected from θd and a word wn is selected from Фzn. The probability of the corpus is given as:

and β are corpus level parameters assumed to be sampled once in the process of generating a corpus. The variables zdn and wdn are word-level variables and are also sampled once for each word in each document.

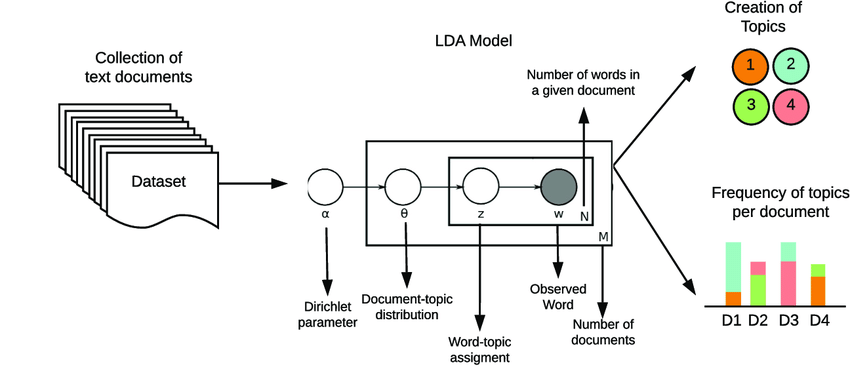


Figure schematic diagram summarizing the process of LDA

***3.2 Latent Semantic Analysis (LSA)***

Also known as latent semantic index (LSI) models the contribution to natural language attributable to combination of words into coherent passages. To construct a semantic space for a language, LSA first casts a large representative text corpus into a rectangular matrix of words by coherent passages, each cell containing a transform of the number of times that a given word appears in each passage. The matrix is then decomposed in such a way that every passage is represented as a vector whose value is the sum of the vectors standing for its component words. Similarities between words and words, passages and words, passages and passages are then computed as dot products, cosines, or other vector-algebraic metrics.

1. **EXPERIMENTAL SETUP**

***Data pre-processing***

First thing I did in this task was to replace unwanted and meaningless terms, for instance “#NAME?”, with space. Afterwards I took care of the empty rows and duplicates in the data set. I also created a new column named “processed”. I removed the stop words which are mostly one, two and three letter words found in the English dictionary. I also removed punctuations numbers then finally lemmatized the words to their basic form. I finally saved this cleaned document into the new column named processed in the data frame. To feed this document into the system, further preprocessing needed to be accomplished. Next, I decomposed the phrases into tokens, that is single words, and included some bi-grams that occurred at least 6 times as tokens then saved in a dictionary to keep track of this vocabularies.

**== Pros(positive reviews) ==**

**Unique number of words before filtering: 2285**

**Unique words after filtering: 894**

**== Cons(negative reviews) ==**

**Unique number of words before filtering 4192**

**Unique words after filtering 1648**

I also weighed the importance of the words in a documents by using the term frequency-inverse term frequency technique which is also a dependency packages in the genism.models.TFidfModel library. I actually tested this two topic modelling method on both the pros and cons separately. So I would just go straight to the point in sharing their respective metrics and choice of best methods based on the c\_v scores.

**5.0 RESULT DISCUSSIONS**

***Latent Dirichlet allocation (PROS)***

I fit the model into the “pros” data and printed the coherence score for 12 topics.

|  |  |
| --- | --- |
| k | Coherence score |
| 3 | 0.276 |
| 4 | 0.263 |
| 5 | 0.263 |
| 6 | 0.271 |
| 7 | ***0.293*** |
| 8 | 0.274 |
| 9 | 0.268 |
| 10 | 0.281 |
| 11 | 0.274 |
| 12 | 0.278 |
| 13 | 0.282 |
| 14 | 0.283 |

*Table 1 shows coherence scores of the 12 topics*

Chart, line chart

Description automatically generated

Figure show coherence score vs the number of topics

From figure 5 above and table 1, the number of topics with the highest coherence score is 7 with coherence score of 0.293. this explains the quality or how well the topics are extracted. The metric measure uses the sliding window, a one-set segmentation of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and finally the cosine similarity.

***Latent Semantic Analysis (PROS)***

K coherence score

Topic 3 : 0.31021621321706944

Topic 4 : 0.3261376819104972

Topic 5 : 0.32245155098602474

**Topic 6 : 0.34598172632664764**

Topic 7 : 0.32779949870401115

Topic 8 : 0.3157294963396493

Topic 9 : 0.3165592660965933

Topic 10 : 0.3202905311573755

Topic 11 : 0.340714620928857

Topic 12 : 0.33063271944994294

Topic 13 : 0.3138979852633104

Topic 14 : 0.3022231290960091

In the LSA model result, topic 6 seems to have the highest coherence score of 0.345 which is better but not to far from the result gotten in the LDA result.

Chart, line chart

Description automatically generated

Figure shows coherent score plotted against number of topics

***Latent Dirichlet allocation (cons)***

I passed the processed cons reviews into the LDA model as a corpus\_tfidf and printed the coherence score for k = 3…14.

|  |  |
| --- | --- |
| k | Coherence score |
| Topic 3 | 0.444 |
| Topic 4 | 0.461 |
| Topic 5 | 0.468 |
| Topic 6 | 0.465 |
| Topic 7 | 0.463 |
| **Topic 8** | **0.471** |
| Topic 9 | 0.434 |
| Topic 10 | 0.437 |
| Topic 11 | 0.456 |
| Topic 12 | 0.438 |
| Topic 13 | 0.421 |
| Topic 14 | 0.429 |

*Table 2 shows coherence score of the 12 topics*

Chart, line chart

Description automatically generated

Figure coherence score vs number of topics

Topic with the highest coherence score is 8 with score 0.471.

Chart, bubble chart

Description automatically generated

Figure shows inter-topic distance map and most salient words

In figure 9 we see the topics shown on the left while the words are on the right. Topics closer to each other are more like each other while those apart are less similar . Chart, bubble chart

Description automatically generated

Figure shows the inter-topi distance map and the top 30 salient words.

In figure 10, we see the top 30 most relevant terms for topic 6 which is the largest topic. The hyper-parameter λ was set to 0.8 from default value of 1.0.

To get a better understanding of what the eight topics represents, it is important we view the top 5 salient words for each topic.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Topic1 | Topic2 | Topic3 | Topic4 | Topic5 | Topic6 | Topic7 | Topic8 |
| 1 | Cut | layoff | Hard | Bad | Organization | People | Project | keep |
| 2 | Moving | Insurance | Sometimes | Workworn | Morale | Year | None | bonus |
| 3 | Good | Health | Challenging | Hour | Made | Get | engineering | market |
| 4 | Reduced | Career | Cutting | Best | Result | Raise | enough | seem |
| 5 | outsourcing | great | limited | program | advancement | high | requirement | heavy |

*Table 3 chows the top 5 words in each 8 topics*

Table 3 above shows how the model groups the reviews (cons) into various clusters of topics. Eight topics with the top salient words in each topic.

Text

Description automatically generated

A picture containing text

Description automatically generated

Figure shows the word cloud for the eight topics

Dominant topics have been extracted for each review from the staffs of Honeywell manufacturing company. The word cloud above in figure 10 shows the words that are peculiar as instances to the topics. Topic 1 cluster of words are 268, topic 2 has 252, topic 3 has 382 cases, topic 4 has 147 cases, topic 5 has 463 instances, topic 6 has 66 only, topic 7 has 257 cases and topic 8 has 165 instances. I also plotted the relationship between the staff’s title and the count of topics as shown in figure 11.

Chart, bar chart

Description automatically generated

Figure show the amount of topics mentioned in the reviews of various staff title

***Latent Semantic Analysis (cons)***

In this model the topic with the highest coherence score is k = 3 with score of 0.46 as seen in figure 10.

In this case , since the LDA model metric performance is better, I proceeded to carry out more analysis on the LDA result to better understand what the 8 topics are about.

Chart, line chart

Description automatically generated

Figure shows coherence score vs number of topics

**5.0 CONCLUSIONS**

In the first review analysis for the positive comments (pros), using the LDA method we got 0.293 for 7 topics which was really low. While using the LSA method we got a coherence score of 0.345 with 6 topics which is better than the result gotten from the LDA result but not a good result. Of course, we know the higher the coherence scores the better. However, a score of 0.345 can be further improved.

In the second review analysis with the negative comments (cons) from the staffs of Honeywell manufacturing company. We got a better result using the LDA method for topic modelling with the highest coherence score of 0.471 for 8 topics and the LSA coherence score of 0.46 for 3 topics. Conclusively, the document clustering of the LDA is more efficient which indicates a better in-depth analysis and this confirms that the latent Dirichlet allocation method is best suitable analysis for this dataset.

**REFERENCE**

David M. Blei (2003). Latent Dirichlet allocation. Computer Science Division University of California Berkeley, CA 94720, USA. [BLEI@CS.BERKELEY.EDU](mailto:BLEI@CS.BERKELEY.EDU)

Andrew Y. Ng (2003). Latent Dirichlet allocation .Computer Science Department Stanford University Stanford, CA94305,USA. [ANG@CS.STANFORD.EDU](mailto:ANG@CS.STANFORD.EDU)

Michael I. Jordan (2003). Latent Dirichlet allocation Computer Science Division and Department of Statistics University of California Berkeley, CA 94720, USA. JORDAN@CS.BERKELEY.EDU

[Dr. Thomas K Landauer (2008), University of Colorado at Boulder and Pearson Knowledge Technologies](http://www.scholarpedia.org/article/User:Thomas_K_Landauer). ). Latent semantic analysis. Scholarpedia, 3(11):4356. [doi:10.4249/scholarpedia.4356](http://dx.doi.org/10.4249/scholarpedia.4356)

[Dr. Susan Dumais (2008), Microsoft Inc., One Microsoft Way, Redmond WA](http://www.scholarpedia.org/article/User:Susan_Dumais). ). Latent semantic analysis. Scholarpedia, 3(11):4356. [doi:10.4249/scholarpedia.4356](http://dx.doi.org/10.4249/scholarpedia.4356)

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

**PROJECT PROPOSAL FOR 7135CEM (RESIT): MODELLING AND OPTIMIZATION UNDER UNCERTAINTY**

**TITLE**: Topic Modelling of reviews on Glassdoor manufactured by Honeywell.

**DATA SET LINK:** www.kaggle.com/datasets/dhirajnimbalkar/topicmodellinghoneywellglassdoorreviews

**DATA SET DESCRIPTION:** The data contains 2000 rows which are majorly reviews of the Honeywell aerospace from customers as well as staffs of Honeywell manufacturing company.

**PROBLEM DESCRIPTION:** Reviews make a great distinction in how consumers/customers/users feel about a product of a company. It could tell how much they are satisfied or not after the purchase and use of a particular product which could help the manufacturing companies identify ways to improve the quality of their products/services. The problem is how do they understand how the customers really feel based on these reviews recorded as “pros” and “cons”.

**WORK PLAN AND APPROACH:**

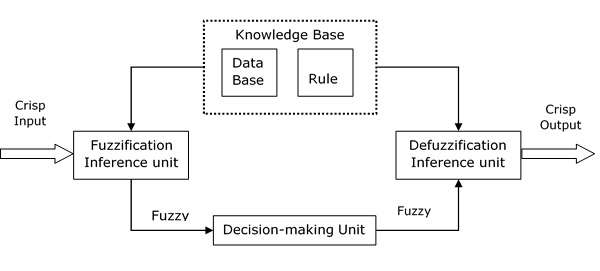
1. Importing necessary dependencies from libraries, loading data, and pre-processing the data which might include removing noise, duplicated rows, unwanted characters/symbols and null spaces.
2. Topic Model selection using the **Latent Semantic Analysis(LSA)** and **Latent Dirichlet Allocation(LDA),** which are both unsupervised machine learning algorithms, for dimension reduction and text classification. LSA takes a matrix of words and decompose them into a separate document-topic matrix and a topic-term matrix. It uses term frequency-inverse document frequency to analyze documents by ranking while preserving the basic statistical relationships of relevance inside a corpus. LDA is a mathematical method for finding the mixture of words that is associated with each topic, while also determining the mixture of topics that describes each document.
3. I would compare both topic modelling methods based on my outcome.

**TASK 2**

**FUZZY LOGIC CONTROLLER DESIGN**

**INTRODUCTION**

The following five functional blocks helps us understand the construction of the fuzzy control system as depicted in the picture found below in figure 14 below. The steps majorly are the input variable, rule inference, aggregation of rule output and finally the defuzzification stage.



In this second task I will be using the MATLAB fuzzy toolbox, Mamdani fuzzy inference system to be precise. Five inputs; temperature, flow of water in pipes, smoke/steam detectors in kitchen, outdoor & indoor light ambience, and 6 outputs; shower temperature, cold & hot valves, extraction fan in kitchen, indoor light and finally the AC/ heater was implemented.

**IMPLEMENTATION OF PART 1**

A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value, or degree of membership in other words, between o and 1. The input space is sometimes referred to as the universe of discourse.

Inputs and membership functions:

1. Diagram

   Description automatically generated Temperature – this input is a sensor that detects the temperature of the house and environs. Its membership functions are shown in the table below

Figure shows membership function plot

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Very cold | -150C to -00C | Triangle |
| Cold | -60C to 50C | Triangle |
| Moderate | 00C to 100C | Triangle |
| Hot | 50C to 150C | Triangle |
| Very hot | 100C to 230C | Triangle |

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Slow | -2 to 0 | Triangle |
| moderate | -1 to 3 | Trapezium |
| Fast | 2 to 4 | Triangle |

1. Flow force of water in pipes Line chart

   Description automatically generated with medium confidence

Figure

1. Smoke/steam detection in kitchen has 5 membership function.

Diagram

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Very low | 0 to 200mg/mm3 | gaussian |
| Low | 150 mg/mm3 to 250 mg/mm3 | gaussian |
| Medium | 100 mg/mm3 to 600 mg/mm3 | gaussian |
| High | 160 mg/mm3 to 800 | gaussian |
| Very high | 700 mg/mm3 to 1800 mg/mm3 | gaussian |

*Smoke/steam membership function*

1. Outdoor light sensor

Diagram

Description automatically generated

Figure show the mf plot for outdoor light

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Very low | 0 to 50 LUX | Trapezium |
| Low | 30 LUX to 80 LUX | Triangle |
| Medium | 50 LUX to 100 LUX | Triangle |
| bright | 80 LUX to 150 LUX | Triangle |
| Very bright | 120 LUX to 250 LUX | Trapezium |

*Smoke/steam membership function*

1. Indoor light ambience has 2 membership functions which are low and high.

Chart, line chart

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Low | 0 to 50 LUX | Triangle |
| High | 50 LUX to 100 LUX | Triangle |

Indoor light ambience

Outputs and membership functions.

1. Shower output variable consist of 3 membership functions.

Diagram, venn diagram

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Cold | 00C to 150C | Triangle |
| Moderate | 100C to 200C | Triangle |
| Hot | 150C to 350C | Triangle |

1. Cold valve

Diagram

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Close fast | -1.5 to -1 | Triangle |
| Close slow | -1 to 0 | Triangle |
| Normal | 0 to 1.5 | Triangle |
| Open slow | 1 to 2 | Triangle |
| Open fast | 1.5 to 2.5 | Triangle |

*Cold valve membership function*

1. Hot valve

A picture containing line chart

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Close fast | -1.5 to -1 | Triangle |
| Close slow | -1 to 0 | Triangle |
| Normal | 0 to 1.5 | Triangle |
| Open slow | 1 to 2 | Triangle |
| Open fast | 1.5 to 2.5 | Triangle |

*hot valve control membership function*

1. Cooker hood extractor has 3 membership functions

Diagram

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Off/low | 0 to 300 rpm | Gaussian |
| Medium | 300 to 600 rpm | Gaussian |
| Fast | 600 to 1200 rpm | Gaussian |

*Cooker extractor control membership function*

1. Indoor light has 5 membership functions

Diagram

Description automatically generated with low confidence

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Very low | 0 to 50 LUX | Gaussian bell |
| Dim | 30 LUX to 80 LUX | Gaussian bell |
| Medium | 50 LUX to 100 LUX | Gaussian bell |
| Bright | 80 LUX to 150 LUX | Gaussian bell |
| Very bright | 120 LUX to 250 LUX | Gaussian bell |

*Indoor light control membership function*

1. AC/heater control output variable

Diagram

Description automatically generated

Figure

|  |  |  |
| --- | --- | --- |
| MF | RANGE | TYPE |
| Cold | -50C to 100C | Gaussian |
| Moderate | 50C to 200C | Gaussian |
| Hot | 100C to 300C | Gaussian |

*AC/heater control membership function*

The fuzzy logic control system on MATLAB toolbox was employed to create the input and outputs respectively

Diagram

Description automatically generated

Fuzzy rules: it contains fuzzy IF-THEN rules.

Interpreting if-then rules is a three-part process. This process is explained in detail in the next section:

1. **Fuzzify inputs**: Resolve all fuzzy statements in the antecedent to a degree of membership between 0 and 1. If there is only one part to the antecedent, then this is the degree of support for the rule.
2. **Apply fuzzy operator to multiple part antecedents**: If there are multiple parts to the antecedent, apply fuzzy logic operators and resolve the antecedent to a single number between 0 and 1. This is the degree of support for the rule.
3. **Apply implication method**: Use the degree of support for the entire rule to shape the output fuzzy set. The consequent of a fuzzy rule assigns an entire fuzzy set to the output. This fuzzy set is represented by a membership function that is chosen to indicate the qualities of the consequent. If the antecedent is only partially true, (i.e., is assigned a value less than 1), then the output fuzzy set is truncated according to the implication method.

In general, one rule alone is not effective. Two or more rules that can play off one another are needed. The output of each rule is a fuzzy set. The output fuzzy sets for each rule are then aggregated into a single output fuzzy set. Finally, the resulting set is defuzzified, or resolved to a single number.

Text

Description automatically generated

Figure show fuzzy rules

Defuzzification: in this unit the fuzzy input is finally converted into crisp output. Centre of gravity (COG), mean of maximum (MOM) and the center average are the 3 examples of defuzzification. I will be using the centroid defuzzification method at this stage. This is seen in figure 27 below. There are 25 rules in total.

I also plotted the surface plot showing the relationship between the inputs and the outputs.

Table

Description automatically generated

Figure show the rules

Surface plots:

Chart, surface chart

Description automatically generated

Figure shows temperature vs flow vs shower

Chart, surface chart

Description automatically generated

Figure shows temperature vs flow vs cold valve

Chart, surface chart

Description automatically generated

Figure shows lamps vs outdoor light vs indoor light ambience

Chart

Description automatically generated

Figure shows cooker hood extractor vs smoke stream

Chart, surface chart

Description automatically generated

Figure shows temperature vs flow vs valve(hot)

Chart, line chart

Description automatically generated

Figure temperature vs AC/heater

Optimization:

Chromosome calculations for inputs

Temperature has 5 MFs = 3+3+3+3+3 = 15

Input flow has 3 MFs = 3+4+3 = 10

Smoke/steam has 5 MFs = 2+2+2+2+2 = 10

Indoor light ambience has 2 MFs = 3+3 =6

Outdoor light has 5 MFs = 4+3+3+3+4 = 17

Total chromosome present in the input is 15+10+10+6+17 = 58 chromosomes

Chromosome calculation for all six outputs

Shower has 3 MFs = 2+2+2 = 6

Cold valve has 5 MFs = 3+3+3+3+3 = 15

Hot valve has 5 MFs as well = 3+3+3+3+3 = 15

Cooker hood extractor has 4 MFs = 2+2+2+2 = 8

Lamps has 5 MFs = 4+4+4+4+4 = 20

AC/heater has 3 MFs = 2+2+2 = 6

Total chromosomes in output population is = 6+15+15+8+20+6 = 68 output chromosomes.

**REFERENCES**

[1] Dubois, Didier, and Henri M. Prade. Fuzzy Sets and Systems: Theory and Applications. Mathematics in Science and Engineering, v. 144. New York: Academic Press, 1980.

[2] Kaufmann, A., and Madan M. Gupta. Introduction to Fuzzy Arithmetic: Theory and Applications. Van Nostrand Reinhold Electrical/Computer Science and Engineering Series. New York, N.Y: Van Nostrand Reinhold Co, 1985.

[3] Lee, C.C. ‘Fuzzy Logic in Control Systems: Fuzzy Logic Controller. I’. IEEE Transactions on Systems, Man, and Cybernetics 20, no. 2 (April 1990): 404–18. <https://doi.org/10.1109/21.52551>.

[4] Lee, C.C. ‘Fuzzy Logic in Control Systems: Fuzzy Logic Controller. II’. IEEE Transactions on Systems, Man, and Cybernetics 20, no. 2 (April 1990): 419–35. <https://doi.org/10.1109/21.52552>.

[5] Mamdani, E.H., and S. Assilian. ‘An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller’. International Journal of Man-Machine Studies 7, no. 1 (January 1975): 1–13.<https://doi.org/10.1016/S0020-7373(75)80002-2>.

[6] Mamdani, E.H. ‘Advances in the Linguistic Synthesis of Fuzzy Controllers’. International Journal of Man-Machine Studies 8, no. 6 (November 1976): 669–78. <https://doi.org/10.1016/S0020-7373(76)80028-4>.

[7] Mamdani, E.H. ‘Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis’. IEEE Transactions on Computers C–26, no. 12 (December 1977): 1182–91. <https://doi.org/10.1109/TC.1977.1674779>.

[8] Schweizer, B. and A. Sklar, 'Associative functions and abstract semi-groups'. Publ. Math. Debrecen 10 (1963): 69–81.

[9] Sugeno, M., "Fuzzy measures and fuzzy integrals: a survey," (M.M. Gupta, G. N. Saridis, and B.R. Gaines, editors) Fuzzy Automata and Decision Processes, pp. 89-102, North-Holland, NY, 1977.

[10] Sugeno, Michio, ed. Industrial Applications of Fuzzy Control. Amsterdam ; New York : New York, N.Y., U.S.A: North-Holland ; Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co, 1985.

[11] Yager, Ronald R. ‘On a General Class of Fuzzy Connectives’. Fuzzy Sets and Systems 4, no. 3 (November 1980): 235–42. <https://doi.org/10.1016/0165-0114(80)90013-5>.

[12] Yager, Ronald R., and Dimitar P. Filev. ‘Generation of Fuzzy Rules by Mountain Clustering’. Journal of Intelligent and Fuzzy Systems 2, no. 3 (1994): 209–19. <https://doi.org/10.3233/IFS-1994-2301>.

[13] Zadeh, L.A. ‘Fuzzy Sets’. Information and Control 8, no. 3 (June 1965): 338–53. <https://doi.org/10.1016/S0019-9958(65)90241-X>.

**APPENDIX**

The codes used in these two tasks can be found in the github repository link below:

[*https://github.com/okonkwou2/modelling-and-optomization-CW-codes/tree/main*](https://github.com/okonkwou2/modelling-and-optomization-CW-codes/tree/main)