Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Grado Matemáticas e Informática

Fuzzy Countries

In this project, a socioeconomic model for various countries is developed using fuzzy logic. The model is implemented in Ciao Prolog with the RFuzzy library and using Python with Scikit-learn to compare the results with real data and assess the model's credibility. Uflese is used for visualizing the results.

Authors: Javier Comyn, Diego Fogued, Francisco J. González
Professor: Susana Muñoz Hernández

Madrid, 2023/2024

Contents

1 Introduction

1.1 Background and motivation for the study

The motivation for this study comes from the need to better understand and model the complex socioeconomic dynamics of different countries. Traditional economic models often can't handle the uncertainty and vagueness in real-world data. Fuzzy logic theory is well-suited for this task, providing a way to deal with these uncertainties. This study aims to create a more accurate and reliable socioeconomic model, which will be compared with real data to ensure its credibility.

At the beginning, we were looking for a project that could be both fascinating and challenging, while also fitting well with the principles of fuzzy logic. We aimed to choose a topic applicable to real life, allowing us to draw conclusions that we might not have realized without applying these tools.

Initially, we considered focusing on psychological analysis or human mental health, as it seemed an interesting application of fuzzy logic. However, we quickly realized that this topic was too broad and complex for our project's scope and that finding reliable data would be difficult.

Instead, we decided to analyze human behavior in a more indirect way by examining the relationship between socio-economic and environmental indicators and the happiness of a country's population. This topic is relevant and interesting because it explores how different aspects of life affect well-being. Moreover, analyzing the happiness of a country's population allows us to compare our results with the World Happiness Report, a well-known study that ranks countries based on happiness levels. This comparison will help us validate our results and assess the credibility of our approach.

1.2 Research objectives

The main objective of this research is to develop a socioeconomic model that provides relevant insights into the economic and environmental conditions of different countries, which traditional models and classical logic cannot achieve. Additionally, the research seeks to use Uflese for visualizing the outcomes, ensuring that the model's findings are both understandable and useful for further analysis. Ultimately, the goal is to establish a credible model that can provide valuable insights into the socioeconomic conditions of various countries that may not be immediately apparent.

To achieve this, we will develop a fuzzy logic system with functions and rules to analyze the relationship between socio-economic and environmental indicators and the happiness of a country's population. We will use data from reputable sources like the World Happiness Report and the World Bank. By comparing the happiness scores we obtain with those in the World Happiness Report, we will validate our results and assess the credibility of our model.

2 Theorical Framework

2.1 Fuzzy Logic

Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1 both inclusive. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. By contrast, in Boolean logic, the truth values of variables may only be the integer values 0 or 1.

Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

3 Methodology

The methodology used in this project can be divided into the following steps:

1. Data Collection: Collecting data from different sources related to socio-economic and environmental indicators.
2. Data Description: Describing the data collected and analyzing its characteristics.
3. Data Preprocessing: Cleaning, transforming, and integrating the data to make it suitable for analysis.
4. Database Design and Development: Designing and developing a database to store the data and integrate it with the fuzzy logic system.
5. Implementation of the Fuzzy System: Developing the fuzzy logic system with functions and rules to model the relationship between the indicators and happiness.
6. Results and Discussion: Presenting and analyzing the results obtained from the fuzzy logic system.
7. Challenges and Solutions: Identifying difficulties encountered during the project and proposing solutions to overcome them.
8. Conclusions and Future Work: Drawing conclusions from the study and suggesting possible future research directions.

4 Database Design and Development

4.1 Data Collection

To gather the data, we used a variety of sources (mainly Kaggle) to obtain information on different socio-economic and environmental indicators for various countries. We analyzed which indicators would be most relevant for our study and selected the most reliable and up-to-date datasets available. Furthermore, we ensured that the data was clean and consistent by performing data cleaning and validation procedures.

4.2 Data Description

The variables in the database include a mix of socio-economic and environmental indicators, which are:

- **high_economic_freedom**: Economic Freedom, measured through indexes such as the Index of Economic Freedom.
- **risk_high_temperature**: Average Surface Temperature, measured in degrees Celsius.
- **alarming_suicide_rate**: Suicide Rate, measured in suicides per 100,000 inhabitants.
- **high_corruption_concern**: Perception of Corruption, measured through corruption perception surveys.
- **dangerous_population_density**: Population Density, measured in people per square kilometer (P/Km^2).
- **huge_agricultural_land_percentage**: Agricultural Land, measured as a percentage of total land area.
- **extensive_surface**: Land Area, measured in square kilometers (Km^2).
- **strong_armed_forces_rate**: Armed Forces Size, measured by the number of active military personnel.
- **high_birth_rate**: Birth Rate, measured in births per 1,000 inhabitants.
- **critical_co2**: CO2 Emissions, measured in metric tons of CO2 per capita.
- **high_cpi_rate**: Consumer Price Index (CPI), measured as an index.
- **high_fertility_rate**: Fertility Rate, measured in births per woman.
- **vast_forested_area_percentage**: Forested Area, measured as a percentage of total land area.

- **wealthy_gdp_per_capita**: Gross Domestic Product (GDP), measured in USD per capita.
- **high_education_primary**: Gross Primary Education Enrollment, measured as a percentage of the relevant age group.
- **high_education_tertiary**: Gross Tertiary Education Enrollment, measured as a percentage of the relevant age group.
- **high_infant_mortality_rate**: Infant Mortality Rate, measured in deaths of infants under one year old per 1,000 live births.
- **long_life_expectancy**: Life Expectancy, measured in years.
- **big_population_size**: Population Size, measured in number of inhabitants.
- **numerous_active_workers**: Labor Force Participation, measured as a percentage of the working-age population.
- **high_tax_revenue_percentage**: Tax Revenue, measured as a percentage of GDP.
- **significant_population_unemployed**: Unemployment Rate, measured as a percentage of the labor force.
- **large_urban_population**: Urban Population, measured as a percentage of the total population.
- **abundant_renewable_energy**: Renewables, measured as a percentage of equivalent primary energy.
- **high_min_wage**: Minimum Wage, measured in USD per month.
- **high_median_age**: Median Age, measured in years.

4.3 Data Preprocessing

Before integrating the data into the database, we performed several preprocessing steps to clean and transform the data. This included handling missing values, normalizing the data, and converting categorical variables into numerical values. Additionally, the different datasets were merged and integrated into a single database, ensuring that the data was consistent and ready for analysis.

5 Data Analysis

The data analysis involved several key steps to ensure that the data was accurately processed and insightful trends were identified.

Initially, we performed exploratory data analysis using the Uflese environment. This allowed us to make several consults and evaluate the results based on our logic and common knowledge.

Outliers were identified and managed through visual inspection and analysis. By carefully reviewing these data points, we ensured that they did not adversely affect the overall analysis.

After defining the fuzzy logic functions, we adjusted them to produce more plausible results. This iterative process involved tweaking the functions and rules based on the observed outputs and, mostly, arbitrary decisions.

Our analysis revealed several significant trends:

- Countries like Iceland, Norway, and Denmark consistently showed logical results for being environmentally friendly and politically stable.
- Contrastingly, countries such as South Africa, Japan, and Thailand were flagged for certain unexpected outcomes, warranting further investigation.
- "Developed Country" category highlighted the importance of factors beyond economic performance, including life expectancy and infant mortality rates, which led to Japan and Spain ranking highly.
- "Economically Stable Countries" presented interesting results, with Thailand appearing at the top due to its very low unemployment rate, despite other economic challenges.

These findings highlight the complex interplay between various socio-economic and environmental indicators and their impact on overall national well-being.

5.1 Implementation of the Fuzzy System

The fuzzy system was developed using the Ciao Prolog environment along with the rfuzzy library. The rfuzzy library provided essential tools for managing fuzzy logic operations and rules.

The fuzzy rules and functions were defined through arbitrary decisions based on our logic and common knowledge. This included categorizing indicators such as GDP per capita, life expectancy, and educational enrollment into fuzzy sets like low, medium, and high.

Integration of the fuzzy logic system with our database was facilitated by the tools provided by rfuzzy. This allowed seamless data retrieval and processing, enabling real-time analysis of the data inputs.

One major challenge was that rfuzzy does not support decimals. To overcome this, we multiplied variables by powers of 10, ensuring that all calculations were performed with integer values. This workaround maintained the accuracy of our results while adhering to the limitations of the rfuzzy library.

5.2 Results and Discussion

The results obtained from the fuzzy logic system were analyzed and compared with traditional methods. This section presents these results and discusses their implications.

The fuzzy logic system provided happiness scores and other insights that closely matched those in the World Happiness Report and other reputable sources. For instance:

- **Clean Country**: Logical results included Iceland, Norway, and Denmark, while unexpected results such as Japan in the lower ranks prompted further investigation into specific factors like pollution levels.
- **Environmentally Friendly Country**: Highlighted countries with substantial forest and agricultural land, such as Brazil, Canada, and Colombia. The contrast in Spain's ranking between this category and the 'Clean Country' category offered intriguing insights into different environmental aspects.
- **Developed Country**: Japan ranked first due to high life expectancy and low infant mortality, demonstrating the multifaceted nature of development beyond just economic metrics.
- **Economically Stable Country**: Thailand's top position, driven by low unemployment, highlighted the nuanced understanding required when interpreting economic stability.

Compared to traditional regression models, the fuzzy logic system demonstrated superior capability in handling uncertainty and vagueness in the data. This led to more nuanced and realistic insights that traditional methods might overlook.

Specific case studies, such as the high ranking of Scandinavian countries across multiple categories, confirmed the system's effectiveness. Conversely, the anomalous ranking of Japan and certain other countries underscored areas for further refinement and investigation.

6 Optimization

6.1 Credibility Calculation

In this section, I will explain how we determined and automated the calculations of credibility for fuzzy functions.

Firstly, we defined two simple algorithms in Python: one to normalize data, and another to compare two sets of normalized data (knowing that one contains normalized real data and the other contains the truth values of our fuzzy functions) to see how the truth values deviate from the real data using MAE (Mean Absolute Error), a common metric used in machine learning.

Secondly, we faced a significant problem: we needed to have the real data in the correct format for the normalization algorithm to work correctly. This led us to apply various transformations to our CSV file using the Pandas library in Python.

Additionally, there were too many fuzzy functions, making it tedious to manually process all the queries. Moreover, if there were changes in the functions or the database, all that work would be futile. Therefore, given that we couldn't find another way to automate the queries, we implemented a program in C that executed the Ciao interpreter in a process, passed the queries through standard input, and collected and processed the query results through standard output.

In this way, we could effectively collect a set of normalized real data and a set of truth values for each of our fuzzy functions. What remained was trivial: to create a Python script that used the implementations we had previously developed to gather all the results into a text file that we could easily consult.

In summary, we applied a criterion such as MAE to compare the results of our fuzzy functions with real values. This way, we obtained credibility values with a logical basis and possibly a way to test the validity of our results in a given sample space.

7 Project Review and Outcomes

7.1 Challenges and Solutions

Throughout the project, several challenges were encountered, ranging from data collection issues to the technical implementation of the fuzzy system.

We aimed to automate the consults using C and Python to calculate credibility values associated with the functions by comparing the results with actual normalized data. Leveraging knowledge from a college subject called Operating Systems and self-documentation, we successfully automated the consults, enhancing the efficiency and accuracy of our evaluations.

Other challenges included ensuring the accuracy and completeness of data from multiple sources, defining appropriate functions and rules, and integrating the fuzzy logic system with the database. These were addressed through strategies such as cross-referencing multiple data sources, using iterative refinement, utilizing robust development tools, and discussing the results with one another.

7.2 Conclusions and Future Work

This study successfully developed and validated a fuzzy logic-based socioeconomic model to analyze various indicators associated to several countries.

The fuzzy logic system proved effective in handling the complexity and uncertainty inherent in socio-economic data. The model demonstrated high accuracy in predicting happiness scores and other indicators, supporting the hypothesis that socio-economic and environmental factors significantly impact well-being. The nuanced insights provided by the fuzzy logic system highlight its potential as a powerful tool for socio-economic analysis.

Despite its success, the model has certain limitations. The reliance on high-quality data from multiple sources means that any inconsistencies or inaccuracies in the data can affect the results. Additionally, defining universally applicable fuzzy rules and functions remains a challenge, as socio-economic conditions vary widely across different regions and cultures.

Future research could focus on expanding the model to include more diverse indicators, such as cultural factors and governance quality. Testing the model's applicability across different regions and cultures would also be beneficial. Furthermore, integrating machine learning techniques with fuzzy logic could enhance the model's predictive capabilities and robustness. This hybrid approach could provide even deeper insights into the complex interplay of factors affecting national well-being. Additionally, the automation of credibility calculations could be extrapolated to design methods that could help in modeling more accurate fuzzy functions, potentially employing more different criteria in addition to MAE.

References

[1] Elgiriyewithana, Nidula, *Global Country Information Dataset 2023.* [Data set]. (2023, 8 julio). Kaggle.
https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023

[2] Hossainds, Belayet, *Renewable Energy world Wide : 1965 2022.* [Data set]. (2023, 3 marzo). Kaggle.
https://www.kaggle.com/datasets/belayethossainds/renewable-energy-world-wide-19652022

[3] Pei Pei Chen, *Minimum wage by country.* [Data set]. (2020, 27 diciembre). Kaggle.
https://www.kaggle.com/datasets/peipeichen/minimum-wage-by-country

[4] My Koryto, *countryinfo.* [Data set]. (2020, 14 abril). Kaggle.
https://www.kaggle.com/datasets/koryto/countryinfo

[5] Fraser Institute, *Economic Freedom of the World.* [Data set]. (2021).
https://www.fraserinstitute.org/economic-freedom/dataset?geozone=world&min-year=2&max-year=0&filter=0&page=dataset&year=2021

[6] Palinatx, *Mean temperature for countries by year 1901-2022.* [Data set]. (2024, 21 marzo). Kaggle.
https://www.kaggle.com/datasets/palinatx/mean-temperature-for-countries-by-year-2014-2022/suggestions?status=pending&yourSuggestions=true

[7] World Health Organization, *Indicators.* [Data set]. (n.d.). World Health Organization.
https://data.who.int/es/indicators

[8] Beach, J., *World Happiness Report 2013-2023* [Data set]. (2023). Kaggle.
https://www.kaggle.com/datasets/joebeachcapital/world-happiness-report-2013-2023

[9] Singh, A. P., *World Happiness Report 2021* [Notebook]. (2021). Kaggle.
https://www.kaggle.com/code/ajaypalsinghlo/world-happiness-report-2021-world/notebook