

Fuzzy Countries

Javier Comyn, Diego Fogued y Francisco J. González

Universidad Politécnica de Madrid

Curso 2023/2024

Índice de contenidos

- 1 Introducción
- 2 Marco teórico
- 3 Metodología
- 4 Diseño de la Base de Datos
- 5 Análisis de Datos
- 6 Optimización del modelo
- 7 Conclusiones y Resultados

Introducción

Background y Motivación

- Motivación: Mejorar la comprensión y modelado de las complejas dinámicas socioeconómicas de los países.
- Problema: Los modelos económicos tradicionales tienen dificultades con las incertidumbres y vaguedades de los datos reales.
- Solución: Utilizar la teoría de la lógica difusa para manejar estas incertidumbres.
- Objetivo: Crear un modelo socioeconómico más preciso y fiable, compararlo con datos reales para garantizar su credibilidad.
- Selección del Proyecto:
 - Criterios: Fascinante, desafiante, adecuado a los principios de la lógica difusa, aplicable a la vida real.
 - Consideración Inicial: Análisis psicológico o salud mental humana.
 - Elección Final: Analizar la relación entre indicadores socioeconómicos y ambientales y la felicidad de la población.

Objetivos del Proyecto

- Objetivo Principal: Desarrollar un modelo socioeconómico que proporcione información sobre las condiciones económicas y ambientales.
- Visualización: Usar Uflese para visualizar los resultados y asegurar su comprensibilidad y utilidad.
- Credibilidad: Establecer un modelo creíble que ofrezca información valiosa.
- Enfoque:
 - Desarrollar un sistema de lógica difusa con funciones y reglas.
 - Analizar la relación entre indicadores socioeconómicos/ambientales y la felicidad de la población.
 - Utilizar datos de fuentes reputadas como el Informe Mundial de la Felicidad y el Banco Mundial.
 - Validar los resultados comparando las puntuaciones de felicidad obtenidas con las del Informe Mundial de la Felicidad.

Marco teórico

- Definición: Lógica multievaluada donde los valores de verdad oscilan entre 0 y 1.
- Propósito: Manejar el concepto de *fuzzy*, a diferencia de la lógica booleana con valores solo de 0 o 1.
- Extensión: Gestionar grados de verdad parcial con funciones específicas para variables lingüísticas.

Metodología

- **Recolección de Datos**
 - Fuentes
- **Descripción de los Datos**
 - Análisis de las características de los datos recolectados.
- **Preprocesamiento de Datos**
- **Diseño y Desarrollo de la Base de Datos**
- **Implementación del Sistema Difuso**
- **Resultados y Discusión**
- **Desafíos y Soluciones**
- **Conclusiones y Trabajo Futuro**

Diseño de la Base de Datos

- Consultar fuentes: Banco Mundial, OMS, Kaggle...
- Escoger indicadores más relevantes para un análisis socioeconómico.
- Elección de los conjuntos de datos más confiables y actualizados.
- Asegurarse de la consistencia y veracidad.

Descripción

- Índice de libertad económica
- Temperatura media ($^{\circ}\text{C}$)
- Tasa suicidios por 100.000 habitantes
- Percepción de la corrupción
- Densidad de población
- Porcentaje de terreno agrícola
- Superficie
- Tamaño del ejército
- Tasa de natalidad
- CO2
- Índice de Precios al Consumidor (IPC)
- Tasa de fertilidad
- Porcentaje de área forestal

Descripción

- PIB per cápita
- Alumnos en educación primaria
- Alumnos en educación post-obligatoria
- Mortalidad infantil
- Esperanza de vida
- Tamaño de la población
- Población activa
- Ingresos fiscales (% del PIB)
- Tasa de paro
- Población urbana
- Energías renovables
- Salario mínimo
- Edad media

Preprocesamiento y Limpieza

- Integrar todas las variables en una única base de datos
- Eliminar inconsistencias
- Tratar valores faltantes
- Convertir todos los valores en enteros

Análisis de Datos

Funciones Difusas

Para cada variable, usando la librería Rfuzzy para Ciao Prolog, definimos una función difusa basándonos en nuestro propio criterio. Los valores se van ajustando progresivamente al ir probando y revisando cada una de ellas.

```
critical_co2(country) :~ function(co2_emissions(country),  
[(0,0),(2000,0.1),(50000,0.3),(100000,0.45),(200000,0.6),(300000,0.8),(1300000,1)]).
```

```
long_life_expectancy(country) :~ function(life_expectancy(country),  
[(350,0), (400,0.2), (550,0.4), (600,0.6), (750,0.8), (900,1)]).
```


Definimos reglas que nos permiten relacionar las distintas funciones, logrando modelizar un sistema complejo de conocimientos socioeconómicos.

Inicialmente, establecemos credibilidad 1 en todas ellas.

```
developed_country(country) :~ rule(mean, ((wealthy_gdp_per_capita(country)), (long_life_expectancy(country)),  
fnot((high_infant_mortality_rate(country))), (high_economic_freedom(country)))) with_credibility (min, 1).
```

```
clean_country(country) :~ rule(mean, ( abundant_renewable_energy(country),  
fnot(critical_co2(country)) )) with_credibility (min, 1).
```

Consultas

Your query:

I'm looking for a

 [show options](#)

10 best results

Results over 70%

Results over 50%

Results over 0%

All results

country	country number	country name	economic freedom index	surface temperature	suicides	people percive corruption	population density	agricultural land	land area	armed forces size	birth	co2 emissions
nº.1	27	Japan	779	1245	2403497	638	347	123	377944	261000	740	1135886
nº.2	19	Germany	773	949	1286067	460	240	477	357022	180000	950	727973
nº.3	18	France	740	1158	1863033	571	119	524	643801	307000	1130	303276
nº.4	26	Italy	725	1352	673713	866	206	432	301340	347000	730	320411
nº.5	10	Canada	798	370	1125687	415	4	69	9984670	72000	1010	544894
nº.6	2	Australia	805	2205	1065003	442	3	482	7741220	58000	1260	375908
nº.7	43	Spain	752	1431	753647	745	94	526	505370	196000	790	244002
nº.8	9	Belgium	750	8553	160000	750	85	330	3040770	730000	1200	160000

Consultas

Your query:

I'm looking for a


show options

Search

10 best results

Results over 70%

Results over 50%

Results over 0%

All results

country	country number	country name	economic freedom index	surface temperature	suicides	people percive corruption	population density	agricultural land	land area	armed forces size	birth	emi
nº.1	22	Iceland	793	243	1244790	673	3	187	103000	0	1200	2
nº.2	34	Norway	767	211	1278203	270	15	27	323802	23000	1040	4
nº.3	14	Denmark	810	896	1474940	179	137	620	43094	15000	1060	3
nº.4	33	New Zealand	843	1118	1317157	242	18	405	268838	9000	1198	3
nº.5	13	Croatia	725	1219	1846353	939	73	276	56594	18000	900	1

Consultas

Your query:

I'm looking for a



Search

10 best results

Results over 70%

Results over 50%

Results over 0%

All results

country	country number	country name	economic freedom index	surface temperature	suicides	people percive corruption	population density	agricultural land	land area	armed forces size	birth	co2 emissions
nº.1	31	Malaysia	719	2647	443000	839	99	263	329847	136000	1675	248289
nº.2	8	Brazil	658	2557	469000	756	25	339	8515770	730000	1392	462299
nº.3	23	Indonesia	693	2601	289667	867	151	315	1904569	676000	1807	563325
nº.4	12	Colombia	660	2497	417000	841	46	403	1138910	481000	1488	97814
nº.5	47	Thailand	707	2690	892977	895	137	433	513120	455000	1034	283763
nº.6	44	Sri Lanka	613	2740	1838990	863	341	437	65610	317000	1583	23362
nº.7	15	Ecuador	646	2142	1001277	843	71	222	283561	41000	1972	41155
nº.8	37	Philippines	701	2670	197800	742	368	417	300000	153000	2055	122287

- **Clean Country:** Resultados lógicos :Islandia, Noruega, Dinamarca... Fue inesperado ver a Japón en los puestos más bajos, descubrimos que era por el CO2.
- **Developed Country:** España por encima de economías mejores, demostrando que no sólo eso define el desarrollo de un país.
- **Environmentally Friendly Country:** Se destacaron países con grandes áreas forestales y agrícolas, como Brasil, Canadá y Colombia.
- **Economically Stable Country:** Tailandia en primera posición, por su bajísimo desempleo. Comprendemos la importancia de la interpretación humana de los resultados.

Comparado con modelos tradicionales, la lógica difusa permite hacer interpretaciones más matizadas y completas de la realidad.

Optimización del modelo

Introducción

En esta sección se explica cómo se determinaron y automatizaron los cálculos de credibilidad para funciones difusas.

- Algoritmos en Python para normalizar datos.
- Comparación de conjuntos de datos normalizados utilizando MAE (Error Absoluto Medio).
- Automatización de consultas en Ciao Prolog

Problemas

- Necesidad de tener datos reales en el formato correcto para el algoritmo de normalización.
- Obtener y procesar resultados de las consultas.

Soluciones

- Aplicación de transformaciones a archivos CSV usando Pandas.
- Implementación de un programa en C que ejecuta el intérprete Ciao.

Objetivo

Crear un programa en C que ejecute el intérprete de Ciao Prolog, automatizando las consultas para las funciones difusas.

- Ejecución del intérprete Ciao Prolog desde un programa en C.
- Envío de consultas a través de la entrada estándar.
- Recopilación y procesamiento de resultados a través de la salida estándar.

Resultado

Automatización completa de las consultas y recolección de datos para el análisis de credibilidad.

Objetivo

Automatizar todo el proceso de consultas y cálculo de credibilidades.

- Recolección de datos normalizados y valores de verdad para funciones difusas.
- Script en Python para consolidar resultados en un archivo de texto.

Resultado

Comparación de valores de verdad con valores reales para obtener valores de credibilidad de forma automática.

Conclusiones y Resultados

Desafíos

- Recopilación y precisión de datos de múltiples fuentes.
- Definición de funciones y reglas apropiadas.
- Integración del sistema de lógica difusa con la base de datos.

Soluciones

- Referencia cruzada de fuentes.
- Refinamiento iterativo.
- Uso de herramientas robustas y discusiones en equipo.

Éxitos

- Desarrollo y validación de un modelo socioeconómico basado en lógica difusa.
- Precisión considerable en las predicciones dadas por los indicadores.

Limitaciones

- Dependencia de datos de alta calidad.
- Definición de reglas difusas universalmente aplicables.

Mejoras Propuestas

- Expandir el modelo para incluir más indicadores diversos.
- Aplicar el modelo en diferentes regiones y culturas.
- Integrar técnicas de aprendizaje automático con lógica difusa.

Potencial

- Extrapolar la automatización de cálculos de credibilidad para diseñar métodos más precisos en el modelado de funciones difusas.