

Ud2 - Resolució de Problemes per Mitjà de Cerques

CE Models d'intel·ligència artificial

Curs 2024-2025

Sebastian Ciscar



IES Jaume II El Just
Tavernes de la Vall d'Alba
Departament d'Informàtica

Continguts

1	Exercici: Creació d'un Sistema de Recomanació de Productes Basat en Distàncies	2
1.1	Passos a seguir:	2
1.2	Objectius de l'exercici:	3
2	Solució de l'Exercici: Creació d'un Sistema de Recomanació per a Ordinadors Portàtils, Ordinadors de Sobretaula i Tauletes	3
2.1	1. Recollida de Dades	3
2.2	2. Preprocessament de Dades	4
2.2.1	2.1 Codificació de Variables Categòriques	4
2.2.2	2.2 Normalització de Dades	5
2.3	3. Càlcul de Distàncies	6
2.4	4. Sistema de Recomanació	7
2.5	5. Visualització (Opcional)	8
2.6	6. Avaluació	8
2.7	Conclusió General	9
2.8	Fórmula de la Distància Euclidiana	9

1 Exercici: Creació d'un Sistema de Recomanació de Productes Basat en Distàncies

Se't demana desenvolupar un sistema de recomanació de productes utilitzant la distància entre productes basada en les seves característiques.

1.1 Passos a seguir:

1. Recollida de dades:

- Genera un conjunt de dades que conté una llista de productes, cadascun amb les seves característiques. Per exemple:
 - *ID del producte*
 - *Preu*
 - *Pes*
 - *Color* (codificat numèricament)
 - *Marca* (codificat numèricament)
 - *Categoria* (codificat numèricament)
 - Altres característiques rellevants.

2. Preprocessament de dades:

- Assegura't que totes les característiques són numèriques. Si hi ha variables categòriques, codifica-les numèricament (per exemple, utilitzant *one-hot encoding* o *label encoding*).
- Normalitza o estandarditza les dades per evitar que certes característiques tinguin més pes en el càlcul de la distància.

3. Càlcul de distàncies:

- Implementa una funció per calcular la distància entre dos productes basant-te en les seves característiques. Pots utilitzar la distància euclidiana, manhattan o qualsevol altra mètrica adequada.

4. Sistema de recomanació:

- Per a un producte donat, troba els n productes més propers (és a dir, amb menor distància) i recomana'ls com a productes similars.
- Prova el teu sistema amb diferents productes i verifica si les recomanacions són coherents.

5. Visualització (opcional):

- Representa gràficament els productes en un espai reduït (per exemple, utilitzant PCA o t-SNE) per visualitzar les similituds entre ells.

6. **Avaluació:**

- Si disposes de dades de preferències d'usuaris o històrics de vendes, avalua l'efectivitat del teu sistema de recomanació comparant les recomanacions amb les eleccions reals dels usuaris.

1.2 Objectius de l'exercici:

- Aplicar tècniques de preprocessament i codificació de dades.
- Entendre i calcular distàncies en espais multidimensionals.
- Desenvolupar un sistema bàsic de recomanació basat en similituds.
- Practicar la visualització i interpretació de dades.

Nota: Aquest exercici et permetrà comprendre com les distàncies en un espai de característiques poden ser utilitzades per recomanar productes similars i com el preprocessament de dades afecta els resultats del sistema de recomanació.

2 Solució de l'Exercici: Creació d'un Sistema de Recomanació per a Ordinadors Portàtils, Ordinadors de Sobretaula i Tauletes

Anem a desenvolupar un sistema de recomanació de productes utilitzant la distància entre ordinadors portàtils, ordinadors de sobretaula i tauletes, basant-nos en les seves característiques.

2.1 1. Recollida de Dades

Crearem un conjunt de dades amb una llista de productes (ordinadors portàtils, de sobretaula i tauletes) i les seves característiques principals.

Conjunt de Dades:

ID	Tipus	Marca	Preu (€)	Processador	RAM (GB)	Emmagatzematge (GB)	Mida Pantalla (")	Pes (kg)
1	Portàtil	Marca A	800	i5	8	256 SSD	15.6	1.8
2	Sobretaula	Marca B	1200	i7	16	512 SSD	24	8
3	Tauleta	Marca C	500	ARM	4	64 Flash	10.1	0.5
4	Portàtil	Marca D	1500	i7	16	1 TB SSD	13.3	1.2
5	Sobretaula	Marca A	700	i3	8	1 TB HDD	21	7
6	Tauleta	Marca B	300	ARM	2	32 Flash	8	0.4
7	Portàtil	Marca C	1000	i5	16	512 SSD	14	2
8	Sobretaula	Marca D	2000	i9	32	2 TB SSD	27	10
9	Tauleta	Marca A	600	ARM	6	128 Flash	12	0.6
10	Portàtil	Marca B	1100	i7	8	256 SSD	15	1.5

2.2 Preprocessament de Dades

2.2.1 Codificació de Variables Categòriques

Codificarem les variables categòriques (Tipus, Marca, Processador) utilitzant *Label Encoding*.

- **Tipus:**

- Portàtil = 0
- Sobretaula = 1
- Tauleta = 2

- **Marca:**

- Marca A = 0
- Marca B = 1
- Marca C = 2
- Marca D = 3

- **Processador:**

- i3 = 0
- i5 = 1
- i7 = 2
- i9 = 3
- ARM = 4

Conjunt de Dades Codificat:

ID	Tipus	Marca	Preu (€)	Processador	RAM	Emmagatzematge	Mida Pantalla	Pes
1	0	0	800	1	8	256	15.6	1.8
2	1	1	1200	2	16	512	24	8
3	2	2	500	4	4	64	10.1	0.5
4	0	3	1500	2	16	1000	13.3	1.2
5	1	0	700	0	8	1000	21	7
6	2	1	300	4	2	32	8	0.4
7	0	2	1000	1	16	512	14	2
8	1	3	2000	3	32	2000	27	10
9	2	0	600	4	6	128	12	0.6
10	0	1	1100	2	8	256	15	1.5

2.2.2 2.2 Normalització de Dades

Per evitar que característiques amb valors més grans influeixin més en el càlcul de la distància, normalitzarem les dades utilitzant la normalització min-max entre 0 i 1.

Càlcul de Min i Max per a cada característica:

- **Tipus:** Min = 0, Max = 2

- **Marca:** Min = 0, Max = 3
- **Preu:** Min = 300, Max = 2000
- **Processador:** Min = 0, Max = 4
- **RAM:** Min = 2, Max = 32
- **Emmagatzematge:** Min = 32, Max = 2000
- **Mida Pantalla:** Min = 8, Max = 27
- **Pes:** Min = 0.4, Max = 10

Aplicació de la Fórmula de Normalització:

$$\text{Valor normalitzat} = \frac{\text{Valor actual} - \text{Min}}{\text{Max} - \text{Min}}$$

Conjunt de Dades Normalitzat:

ID	Tipus	Marca	Preu	Processador	RAM	Emmagatzematge	Mida Pantalla	Pes
1	0.0	0.0	0.294	0.25	0.206	0.112	0.389	0.145
2	0.5	0.333	0.529	0.5	0.451	0.24	0.842	0.771
3	1.0	0.667	0.118	1.0	0.065	0.016	0.116	0.011
4	0.0	1.0	0.706	0.5	0.451	0.486	0.278	0.089
5	0.5	0.0	0.235	0.0	0.206	0.486	0.684	0.725
6	1.0	0.333	0.0	1.0	0.0	0.0	0.0	0.0
7	0.0	0.667	0.412	0.25	0.451	0.24	0.316	0.167
8	0.5	1.0	1.0	0.75	1.0	1.0	1.0	1.0
9	1.0	0.0	0.176	1.0	0.129	0.048	0.211	0.022
10	0.0	0.333	0.471	0.5	0.206	0.112	0.368	0.122

2.3 3. Càlcul de Distàncies

Utilitzarem la **distància Euclidiana** per calcular la similitud entre productes.

Funció de Distància Euclidiana:

$$\text{Distància} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

On (x_i) i (y_i) són les característiques dels productes.

2.4 4. Sistema de Recomanació

Per a cada producte, trobarem els 2 productes més propers (amb menor distància).

Exemple: Recomanacions per al Producte ID 1 (Ordinador Portàtil de Marca A)

Característiques del Producte 1:

Tipus	Marca	Preu	Processador	RAM	Emmagatzematge	Mida Pantalla	Pes
0.0	0.0	0.294	0.25	0.206	0.112	0.389	0.145

Càlcul de la Distància amb la resta de productes:

- **Distància amb Producte 2:**

$$\sqrt{(0.0 - 0.5)^2 + (0.0 - 0.333)^2 + (0.294 - 0.529)^2 + \dots} \approx 1.054$$

- **Distància amb Producte 3:**

$$\sqrt{(0.0 - 1.0)^2 + (0.0 - 0.667)^2 + \dots} \approx 1.583$$

- **Distància amb Producte 4:**

$$\sqrt{(0.0 - 0.0)^2 + (0.0 - 1.0)^2 + \dots} \approx 0.679$$

- **Distància amb Producte 5:**

$$\sqrt{(0.0 - 0.5)^2 + (0.0 - 0.0)^2 + \dots} \approx 0.887$$

- **Distància amb Producte 6:**

$$\sqrt{(0.0 - 1.0)^2 + (0.0 - 0.333)^2 + \dots} \approx 1.663$$

- **Distància amb Producte 7:**

$$\sqrt{(0.0 - 0.0)^2 + (0.0 - 0.667)^2 + \dots} \approx 0.361$$

- **Distància amb Producte 8:**

$$\sqrt{(0.0 - 0.5)^2 + (0.0 - 1.0)^2 + \dots} \approx 1.415$$

- **Distància amb Producte 9:**

$$\sqrt{(0.0 - 1.0)^2 + (0.0 - 0.0)^2 + \dots} \approx 1.532$$

- **Distància amb Producte 10:**

$$\sqrt{(0.0 - 0.0)^2 + (0.0 - 0.333)^2 + \dots} \approx 0.248$$

Ordenació de les Distàncies:

1. **Producte 10:** 0.248
2. **Producte 7:** 0.361
3. **Producte 4:** 0.679

Recomanacions per al Producte 1:

- **1r Recomanat:** Producte 10 (Ordinador Portàtil de Marca B)
- **2n Recomanat:** Producte 7 (Ordinador Portàtil de Marca C)

Repetim el procés per a altres productes si és necessari.

2.5 5. Visualització (Opcional)

Podem utilitzar l'Anàlisi de Components Principals (PCA) per reduir la dimensionalitat i representar els productes en 2D.

2.6 6. Avaluació

Com que no disposem de dades de preferències d'usuaris, podem avaluar subjectivament si les recomanacions tenen sentit.

Anàlisi de les Recomanacions per al Producte 1:

- **Producte 10:** És un ordinador portàtil amb característiques similars (mateix tipus, preu similar, mateix processador).
- **Producte 7:** També és un portàtil amb característiques similars, però d'una marca diferent.

Conclusió: Les recomanacions semblen coherents, ja que els productes recomanats són portàtils amb característiques similars al producte original.

2.7 Conclusió General

Hem desenvolupat un sistema de recomanació bàsic basat en la distància entre productes utilitzant les seves característiques. Hem:

- Creat un conjunt de dades amb ordinadors portàtils, de sobretaula i tauletes.
- Preprocessat les dades codificant variables categòriques i normalitzant les característiques.
- Calculat les distàncies entre productes utilitzant la distància Euclidiana.
- Generat recomanacions basades en els productes més propers en termes de característiques.

Aquest sistema pot ser millorat incorporant més dades, ajustant les mètriques de distància o utilitzant tècniques més avançades de sistemes de recomanació.

2.8 Fórmula de la Distància Euclidiana

Per calcular la distància entre dos productes basant-nos en les seves característiques numèriques, utilitzem la **distància Euclidiana**:

$$\text{Distància}(A, B) = \sqrt{\sum_{i=1}^n (x_{iA} - x_{iB})^2}$$

On:

- (n) és el nombre total de característiques.
- (x_{iA}) és el valor de la característica (i) del producte **A**.
- (x_{iB}) és el valor de la característica (i) del producte **B**.

Aquesta fórmula ens permet quantificar la similitud entre dos productes: com més petita sigui la distància, més similars són els productes en termes de les seves característiques.

Nota: És important assegurar-se que totes les característiques estiguin en la mateixa escala (per exemple, mitjançant la normalització) abans de calcular la distància, per evitar que alguna característica tingui més pes que les altres.