Image-based Recommendations on Styles and Substitutes

Bigdata Mining LAB
2017107684
JinHee Kim

Motivation

- People while shopping often get confused between certain products which one to buy or what other products go by their taste which they are currently viewing.
- Ex) Shoe + Suit Shirt

- Reviews and Purchase History -> "Cold Start" Issue

-> This is solved by modeling visual preference through images.

- Goal: Build a graph-based system to help you buy products that fit people's needs.

Introduct

- Dataset
  - Amazon Web Store based
  - 6 million objects, over 180 million relationships

Category	Users	Items	Ratings	Edges
Books	8,201,127	1,606,219	25,875,237	51,276,522
Cell Phones & Accessories	2,296,534	223,680	5,929,668	4,485,570
Clothing, Shoes & Jewelry	3,260,278	773,465	25,361,968	16,508,162
Digital Music	490,058	91,236	950,621	1,615,473
Electronics	4,248,431	305,029	11,355,142	7,500,100
Grocery & Gourmet Food	774,095	120,774	1,997,599	4,452,989
Home & Kitchen	2,541,693	282,779	6,543,736	9,240,125
Movies & TV	2,114,748	150,334	6,174,098	5,474,976
Musical Instruments	353,983	65,588	596,095	1,719,204
Office Products	919,512	94,820	1,514,235	3,257,651
Toys & Games	1,352,110	259,290	2,386,102	13,921,925
Total	20,980,320	5,933,184	143,663,229	180,827,502

Introduct

- Relationship Type

- users who viewed X also viewed Y
- users who viewed X eventually bought Y

substitute

- users who bought X also bought Y
- users bought X and Y simultaneously

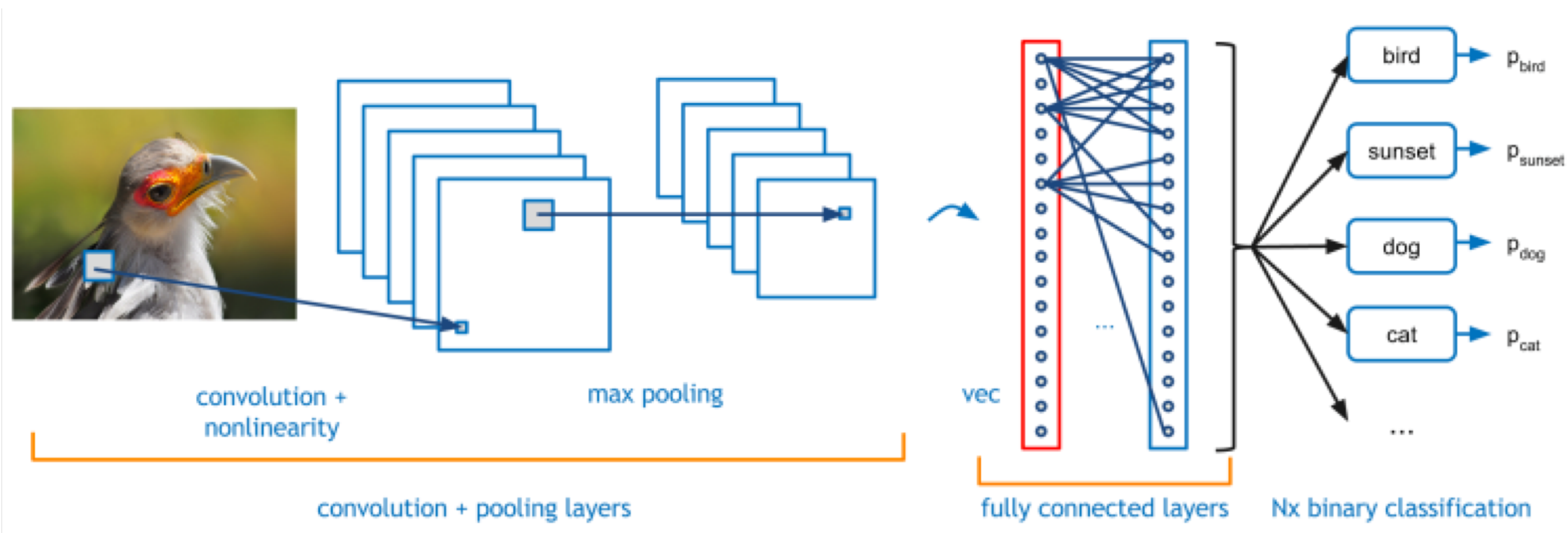
Complementary

➡ Visual approach through images, modeling visual preferences through relationships

MODEL

- **Extract image features**

➢ **Caffe deep learning framework**



MODEL

- **Extract image features**

➢ **CNN(Convolutional neural network)**

- 5 convolutional layer
- 3 fully-connected layer
- ImageNet pre-training
- Using fully-connected 2 layer output (7 layer)

> **F-dimensional feature vector output**

$$\mathbf{x} \in \mathbb{R}^F$$

Feature vector computed from object image

MODEL

- **Relationship sets - learning which products work together**

➢ **Relationships Between Objects**

- Dataset : Includes four types of relationships
- Goal : Learning the parameterized distance transform d(x_i, x_j) so that the feature vector for the relation set ($r_{ij} \in \mathcal{R}$) assigns a lower distance than the relation set($r_{ij} \notin \mathcal{R}$)

• Distance and probability : Use the shift sigmoid function to map the distance and probability

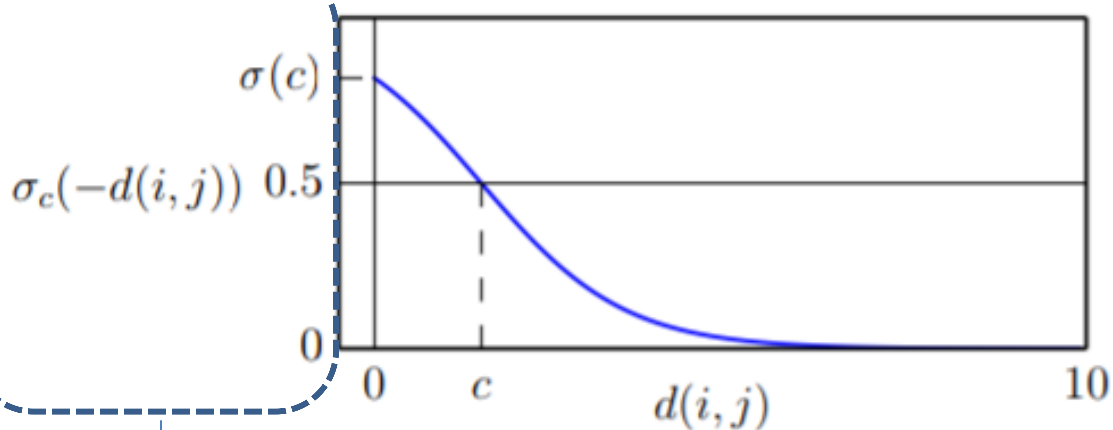
$$P(r_{ij} \in \mathcal{R}) = \sigma_c(-d(\mathbf{x}_i, \mathbf{x}_j)) = \frac{1}{1 + e^{d(\mathbf{x}_i, \mathbf{x}_j) - c}}.$$

Shift sigmoid function with parameter c

Parameterrized distance between x_i and x_j

MODEL

- **Relationship sets**



Relevance

$d(x_i, x_j) = c$ Relevance: 0.5

$d(x_i, x_j) > c$ Probability reduction

$d(x_i, x_j) < c$ Increase probability

MODEL

▪ Relationship sets

➢ Potential distance function

• Weighted nearest neighbor : Learn which feature dimensions are associated with a particular relationship

$$d_{\mathbf{w}}(\mathbf{x}_i,\mathbf{x}_j) = \|\mathbf{w} \circ (\mathbf{x}_i - \mathbf{x}_j)\|_2^2$$

Hadamard product

• Mahalanobis transform : Image Features + Mahalanobis distance

- How different feature dimensions relate to each other
- Weights are defined at the level of pairs of features.

F × F mahalanobis transform matrix

$$d_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)^T$$

Test subset

MODEL

Relationship sets

Potential distance function

- Problem : When you consider the size of your dataset, you have too many parameters.(overfitting)

Ex) F = 2^{12} Features -> Approximately 8 million parameters

- Final object visual + relationship distance function (low-rank transform)

$$d_{\mathbf{Y}}(\mathbf{x}_i,\mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)\mathbf{Y}\mathbf{Y}^T(\mathbf{x}_i - \mathbf{x}_j)^T$$
$$= \|(\mathbf{x}_i - \mathbf{x}_j)\mathbf{Y}\|_2^2.$$

$$\mathbf{M} \simeq \mathbf{Y}\mathbf{Y}^T$$

F × K matrix

MODEL

- **Personalizing styles to individual users**

- Personalize the style dimension that you think is important for individual users

- Customized distance function $d_{Y,u}(x_i, x_j)$ to measure the distance between items i and j according to user u

$$d_{\mathbf{Y},u}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)\mathbf{Y}\mathbf{D}^{(u)}\mathbf{Y}^T(\mathbf{x}_i - \mathbf{x}_j)^T$$

Diagonal user-personalization matrix for user u

Training

- **optimize the log likelihood**

- negative set $\mathcal{Q} = \{r_{i,j} | r_{i,j} \notin \mathcal{R}\}$ randomly selected log likelihood optimization

$$l(\mathbf{Y}, c|\mathcal{R}, \mathcal{Q}) = \sum_{r_{ij} \in \mathcal{R}} \log(\sigma_c(-d_{\mathbf{Y}}(\mathbf{x}_i, \mathbf{x}_j))) + \sum_{r_{ij} \in \mathcal{Q}} \log(1 - \sigma_c(-d_{\mathbf{Y}}(\mathbf{x}_i, \mathbf{x}_j)))$$

- **gradient ascent**

- Optimize $l(Y, c|\mathcal{R}, \mathcal{Q})$ for Y and C achieved with a gradient ascent

- **L-BFGS (Limited-memory BFGS)**

- Used to minimize many variables
- Assuming an arbitrary solution and estimating the next position by using a first derivative value and a second derivative value at this position, it is a method of finding a minimum point.

Experiments - 1

- **Dataset**

- Top-level categories (books, movies, music, etc.)
- Apparel categories are further divided into second level categories (men, women, boys, etc.)

- **Experimental Protocol**

- Each category and graph type forms a single experiment (ex. Predicted 'bought together' relationship for women's clothing)
- Goal: distinguish relationships from non-relationship
- Positive relationships and a random sample of non-relationships of equal size
- All results are reported on the test set

Experiments - 1

Category	method	substitutes: buy after viewing	substitutes: also viewed	complements: also bought	complements: bought together
Books	WNN	66.5%	62.8%	63.3%	65.4%
	$K = 10$	70.1%	68.6%	69.3%	68.1%
	$K = 100$	71.2%	69.8%	71.2%	68.6%
Cell Phones and Accessories	WNN	73.4%	66.4%	69.1%	79.3%
	$K = 10$	84.3%	78.9%	78.7%	83.1%
	$K = 100$	85.9%	83.1%	83.2%	87.7%
Clothing, Shoes, and Jewelry	WNN	·	77.2%	74.2%	78.3%
	$K = 10$	·	87.5%	84.7%	89.7%
	$K = 100$	·	88.8%	88.7%	92.5%
Digital Music	WNN	60.2%	56.7%	62.2%	53.3%
	$K = 10$	68.7%	60.9%	74.7%	56.0%
	$K = 100$	72.3%	63.8%	76.2%	59.0%
Electronics	WNN	76.5%	73.8%	67.6%	73.5%
	$K = 10$	83.6%	80.3%	77.8%	79.6%
	$K = 100$	86.4%	84.0%	82.6%	83.2%
Grocery and Gourmet Food	WNN	·	69.2%	70.7%	68.5%
	$K = 10$	·	77.8%	81.2%	79.6%
	$K = 100$	·	82.5%	85.2%	84.5%
Home and Kitchen	WNN	75.1%	68.3%	70.4%	76.6%
	$K = 10$	78.5%	80.5%	78.8%	79.3%
	$K = 100$	81.6%	83.8%	83.4%	83.2%
Movies and TV	WNN	66.8%	65.6%	61.6%	59.6%
	$K = 10$	71.9%	69.6%	72.8%	67.6%
	$K = 100$	72.3%	70.0%	77.3%	70.7%
Musical Instruments	WNN	79.0%	76.0%	75.0%	77.2%
	$K = 10$	84.7%	87.0%	85.3%	82.3%
	$K = 100$	89.5%	87.2%	84.4%	84.7%
Office Products	WNN	72.8%	75.0%	74.4%	73.7%
	$K = 10$	81.2%	84.0%	84.1%	78.6%
	$K = 100$	85.9%	87.2%	85.8%	80.9%
Toys and Games	WNN	67.0%	72.8%	71.7%	77.6%
	$K = 10$	75.8%	78.3%	78.4%	80.3%
	$K = 100$	77.1%	81.9%	82.4%	82.6%

- Result of selecting the top category

- The learning model has only a reference to the object image.

- The preferences of Substitutes and Complements are almost the same.

Experiments - 1

Category	method	substitutes also viewed	complements also bought	bought together
Baby	CT	77.1%	70.5%	80.1%
	WNN	83.0%	87.7%	81.7%
	$K = 10$	92.2%	92.7%	91.5%
	$K = 100$	94.6%	94.3%	93.3%
Boots	CT	75.0%	72.7%	74.2%
	WNN	83.9%	85.6%	84.7%
	$K = 10$	93.0%	94.9%	95.4%
	$K = 100$	94.6%	96.8%	96.4%
Boys	CT	81.9%	77.3%	83.1%
	WNN	85.0%	87.2%	87.9%
	$K = 10$	94.4%	94.1%	93.8%
	$K = 100$	96.5%	95.8%	95.1%
Girls	CT	83.0%	76.2%	78.7%
	WNN	83.3%	86.0%	84.8%
	$K = 10$	94.5%	93.6%	93.0%
	$K = 100$	96.1%	95.3%	94.5%
Jewelry	CT	50.1%	49.5%	51.1%
	WNN	81.2%	81.6%	75.8%
	$K = 10$	89.6%	89.3%	82.8%
	$K = 100$	89.1%	91.6%	86.4%
Men	CT	88.2%	78.4%	83.6%
	WNN	86.9%	78.4%	82.3%
	$K = 10$	91.6%	89.8%	92.1%
	$K = 100$	92.6%	93.3%	95.1%
Novelty Costumes	CT	79.1%	76.3%	81.5%
	WNN	80.1%	74.1%	76.0%
	$K = 10$	86.3%	86.6%	85.0%
	$K = 100$	89.2%	90.0%	89.1%
Shoes and Accessories	CT	81.3%	78.1%	90.4%
	WNN	75.4%	80.2%	77.9%
	$K = 10$	89.7%	90.4%	93.5%
	$K = 100$	92.3%	94.7%	96.2%
Women	CT	86.8%	79.1%	84.3%
	WNN	78.8%	76.1%	80.0%
	$K = 10$	88.9%	87.8%	91.5%
	$K = 100$	90.4%	91.2%	94.3%

- Results for clothing data

- The learning model has only a reference to the object image.

- The preferences of Substitutes and Complements are almost the same.

Experiments - 2

- **Personalized recommendations**
  - Evaluate the ability to customize the model's joint purchase recommendations to individual users
  - Usage of a tuple data set of the form (i, j, u) for a product i and a pair of products purchased by user u
  - Randomly sampled 50% of joint and non-joint purchases

Category	method	accuracy
Men's clothing	CT	84.8%
	WNN	84.3%
	$K = 10$, no personalization	90.9%
	$K = 10$, personalized	93.2%
Women's clothing	CT	80.5%
	WNN	80.8%
	$K = 10$, no personalization	87.6%
	$K = 10$, personalized	89.1%

User-customized terms yes / no

Visualizing Style Space

- **Visualizing Style Space**
  - K-means clustering in K-dimension

Embedding Boys clothing in 2-dimension

- Sport shoes are made of sandals or slippers.
- Underwear moves gradually toward clothing

Generating recommendations

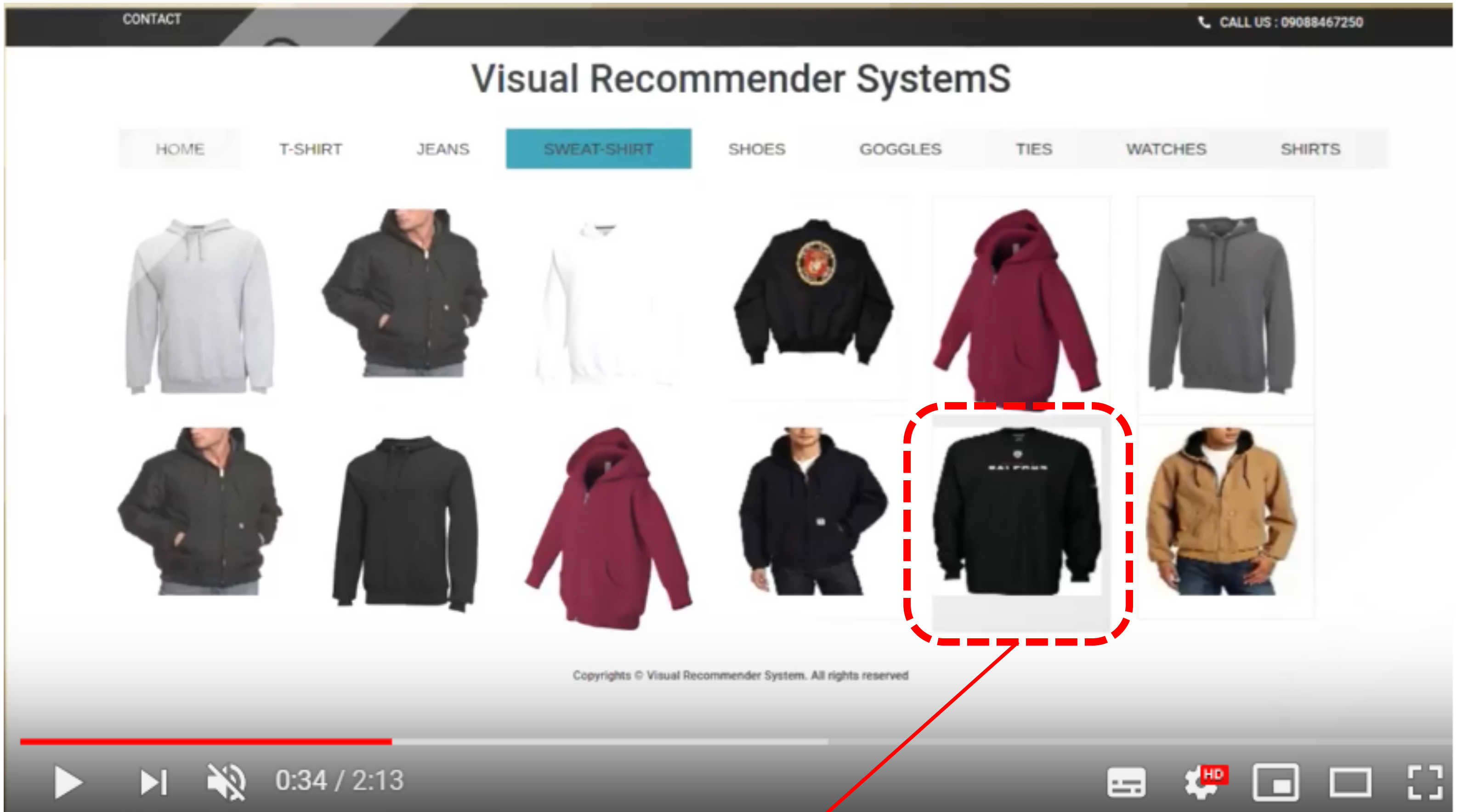
- **Generating recommendations**
  - The proposed model can recommend complementary products.

1) Detailed category hierarchy
-> Definition of 'dress' as a combination of bottoms, tops, shoes, and accessories in the category of women's or men's clothing

2) Given a query item, it indicates items that are likely to be linked based on the visual style.

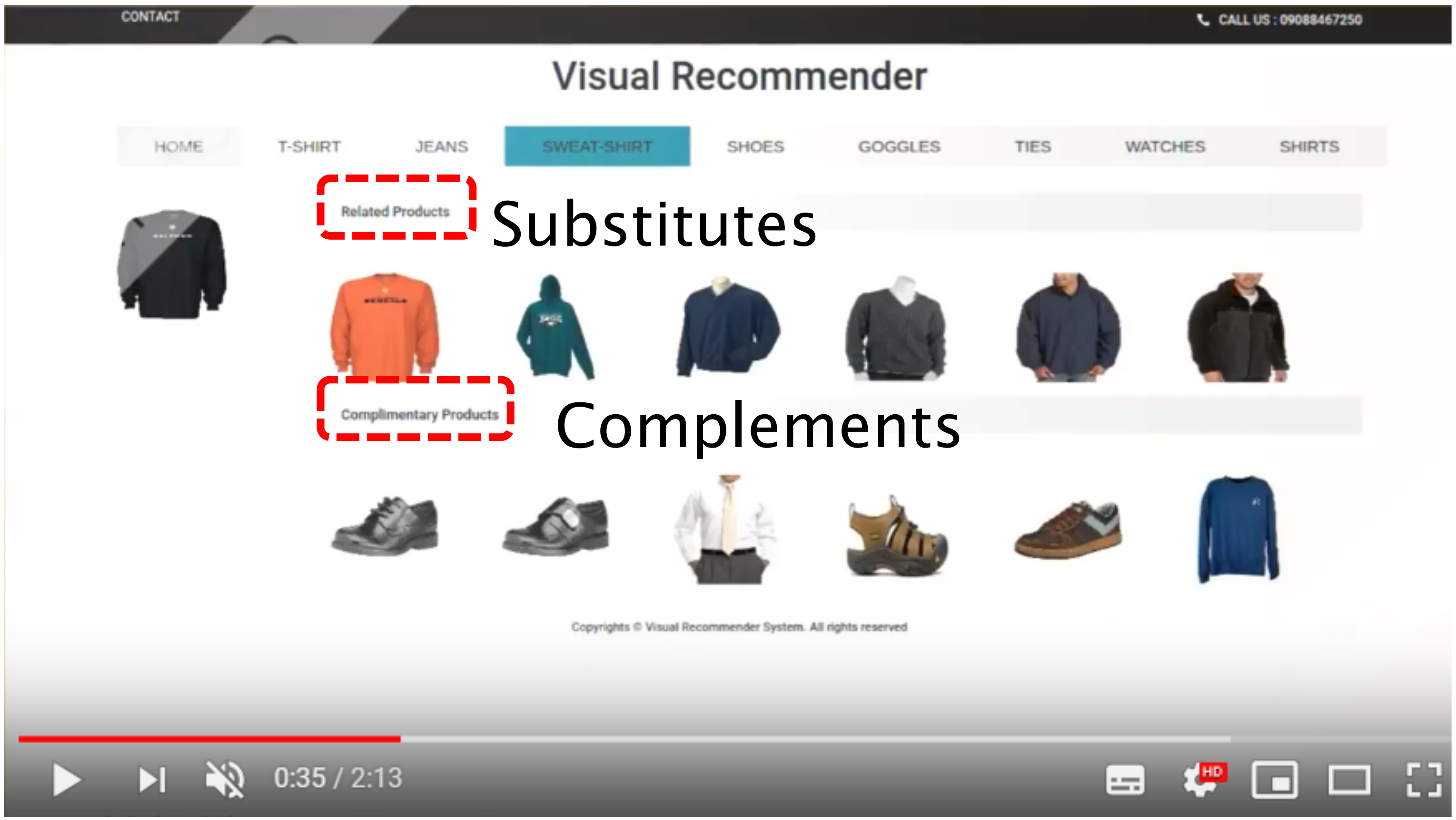
$$\underset{j \in \mathcal{C}}{\operatorname{argmax}}\, P_{\mathbf{Y}}(r_{qj} \in \mathcal{R})$$

Interface



Click : query

Interface



CONTACT
CALL US : 09088467250
Visual Recommender
HOME
T-SHIRT
JEANS
SWEAT-SHIRT
SHOES
GOGGLES
TIES
WATCHES
SHIRTS
Related Products
Substitutes
Complimentary Products
Complements
Copyrights © Visual Recommender System. All rights reserved
0:35 / 2:13
HD

Conclusion

- It is possible to model various visual relationships beyond simply visual similarity.
- create a model that makes items complementary.

Thank you