Feature Selection using Ant Colony Optimization

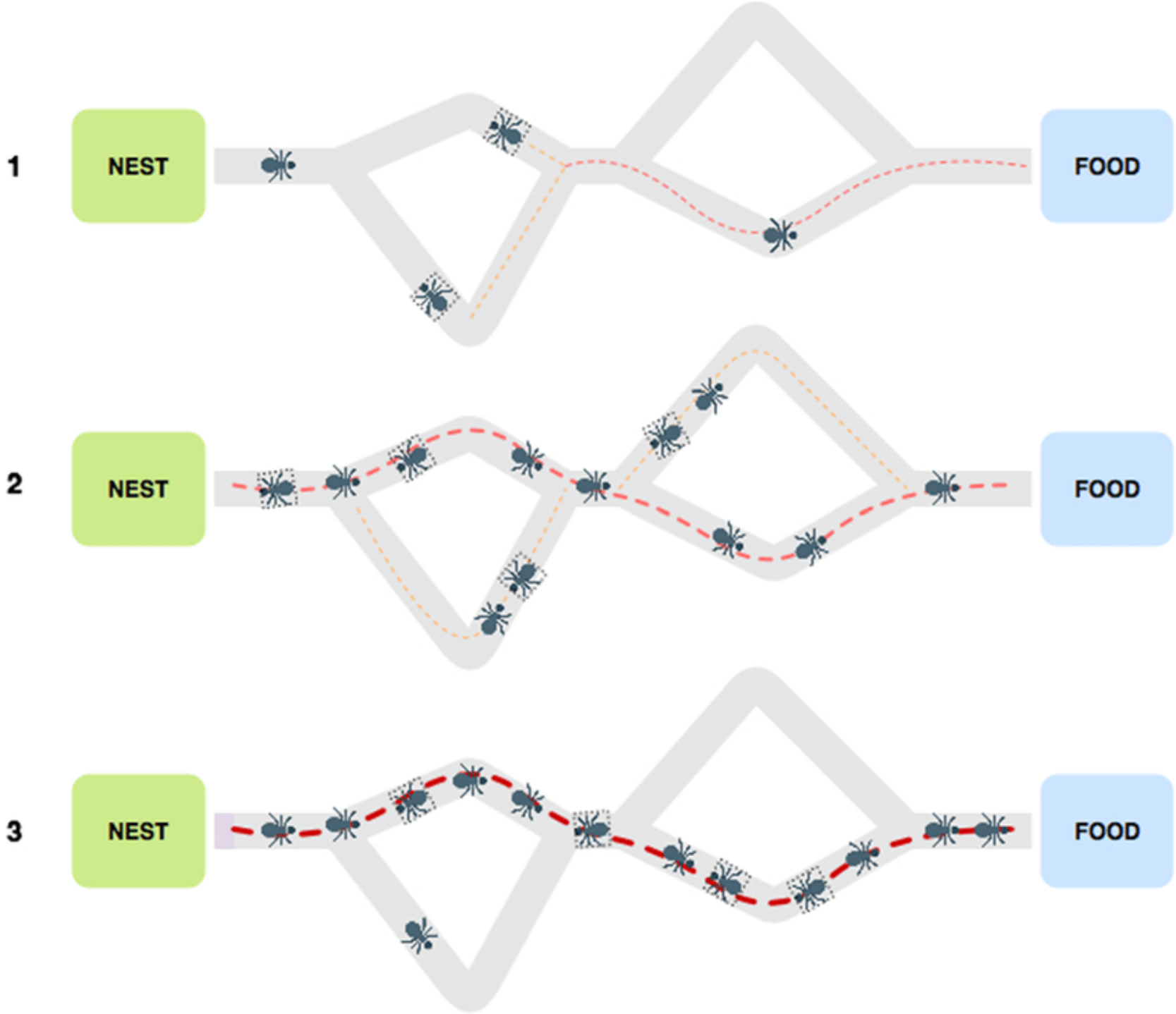
Mert Can BENLİ - Emre KUTLU

Feature Selection

- **Feature selection** is the process of reducing the number of input variables when developing a predictive model.
- It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model.
- It selects the most relevant features.

Ant Colony Algorithm

In the natural world, ants of some species (initially) wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep travelling at random, but instead to follow the trail, returning and reinforcing it if they eventually find food


1
NEST
FOOD
2
NEST
FOOD
3
NEST
FOOD

Ant Colony Algorithm

- Over time, however, the pheromone trail starts to evaporate, thus reducing its attractive strength.
- The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate.
- A short path, by comparison, gets marched over more frequently, and thus the pheromone density becomes higher on shorter paths than longer ones.

General Overview of our System

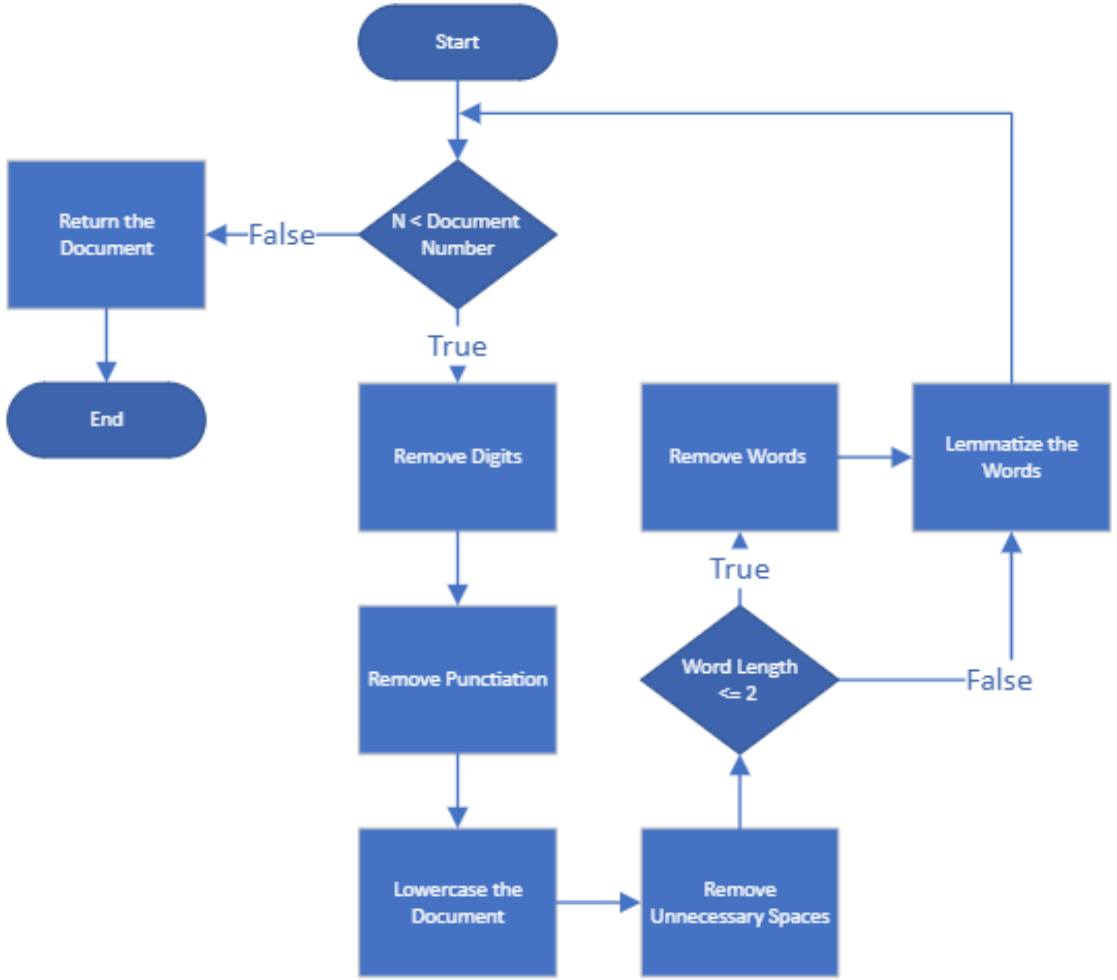
Feature Extraction:

Input : Corpus

Output : Extracted feature set, their term frequencies and TF-IDF weights

Step 1 : Preprocess the document

Preprocess


Start
N < Document Number
False
Return the Document
End
True
Remove Digits
Remove Punctiation
Lowercase the Document
Remove Unnecessary Spaces
Word Length <= 2
True
Remove Words
False
Lemmatize the Words

General Overview of our System

Step 2 : Build BoW and TF-IDF representation

BoW : The bag-of-words model is a simplifying representation used in natural language processing. In this model, a text is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.

TF-IDF : Term Frequency-Inverse Document Frequency is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus

Bag of Words

	she	loves	pizza	is	delicious	a	good	person	people	are	the	best
She loves pizza, pizza is delicious	1	1	2	1	1	0	0	0	0	0	0	0
She is a good person	1	0	0	1	0	1	1	1	0	0	0	0
good people are the best	0	0	0	0	0	0	1	0	1	1	1	1

TF-IDF

Words	TF (for A)	TF (for B)	IDF	TFIDF (A)	TFIDF (B)
Jupiter	1/5	0	ln(2/1) = 0.69	0.138	0
Is	1/5	1/8	ln(2/2) = 0	0	0
The	1/5	2/8	ln(2/2) = 0	0	0
largest	1/5	0	ln(2/1) = 0.69	0.138	0
Planet	1/5	1/8	ln(2/2) = 0	0.138	0
Mars	0	1/8	ln(2/1) = 0.69	0	0.086
Fourth	0	1/8	ln(2/1) = 0.69	0	0.086
From	0	1/8	ln(2/1) = 0.69	0	0.086
Sun	0	1/8	ln(2/1) = 0.69	0	0.086

General Overview of our System

Step 3 : Ant Colony Algorithm

Step 3.1: Initialize Parameters

- Epoch : Maximum iteration number
- Ant : Number of ants
- Feature count : Number of features to select
- $\tau(initial)$: Initial pheromone value
- α : Pheromone exponent value
- β : Heuristic exponent value
- ρ : Pheromone evaporation coefficient

General Overview of our System

Step 3.1:

for each ant:

for each feature:

Select Feature according to the Transition Rule

$$Selection\ Probability = \frac{\tau(feature)^{\alpha} * IDF(feature)^{\beta}}{\sum_{0}^{n} \tau^{\alpha} * IDF^{\beta}}$$

Delete the visited feature from feature list.

Add selected feature to the ant's path.

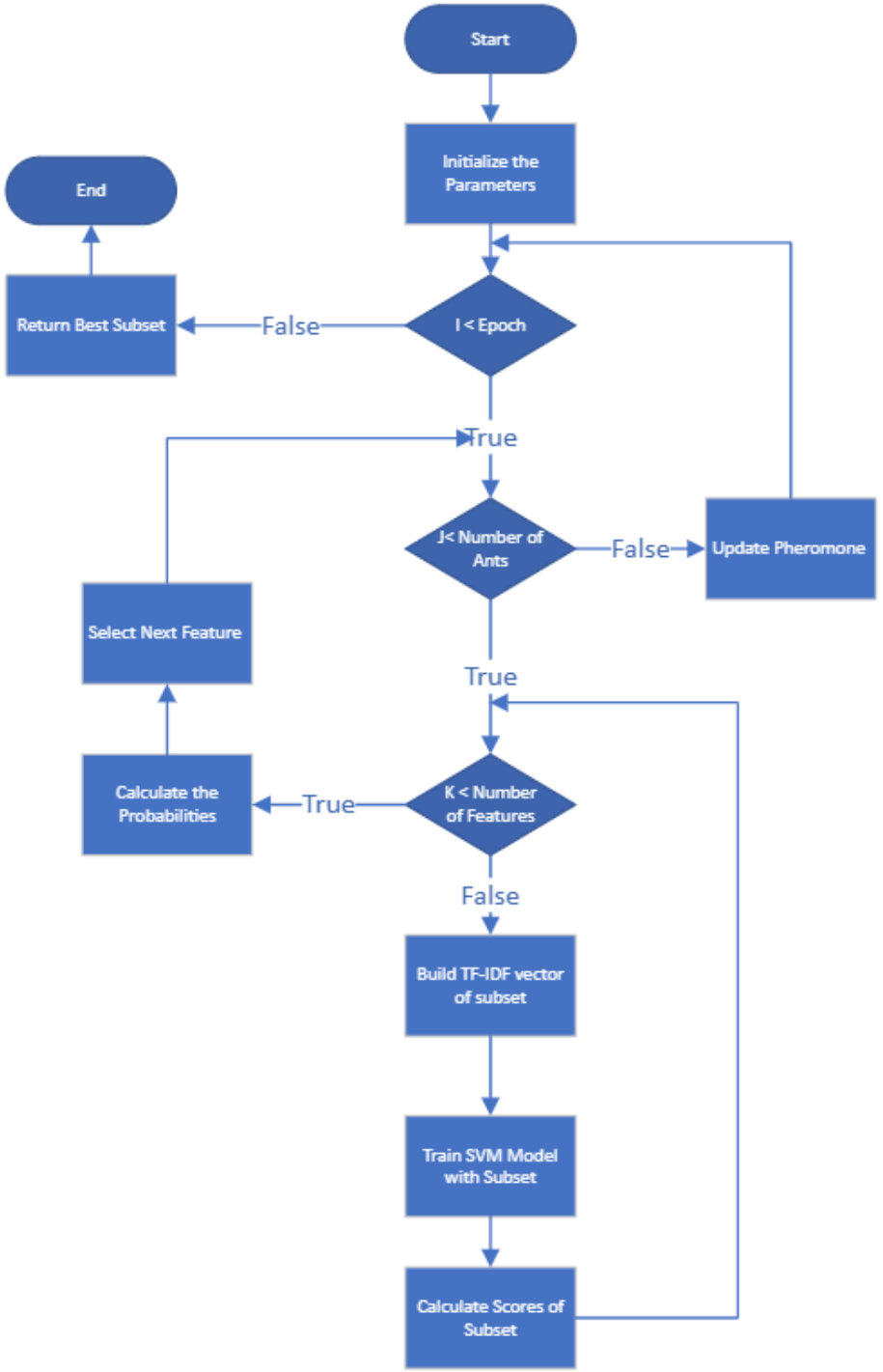
Train the model with constructed subset of ant and calculate F1 scores.

General Overview of our System

Step 3.4: Evaporate and Update Pheromones according to following rules:

$$\text{F}eature\ Pheromone(i) = \sum_{k=0}^{Ant} F1\ Score(k) * \tau(Initial)$$

$$\tau(feature) = \underline{\rho}\tau(feature) + Feature\ Pheromone(i)$$


Start
Initialize the Parameters
I < Epoch
False
Return Best Subset
End
True
J< Number of Ants
False
Update Pheromone
True
K < Number of Features
True
Calculate the Probabilities
Select Next Feature
False
Build TF-IDF vector of subset
Train SVM Model with Subset
Calculate Scores of Subset

Conclusion

- In this work we have implemented ACO based Feature Selection for multilabel text-news classification problem.
- For future works, word embeddings can be used as heuristic value.

Any Questions