INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

Invitation to MTech Thesis Defense of **Biswadeep Khan** July 06, 2018 (Friday): 03.00-04.30 IST

In Partial Fulfillment of the Requirements for the Degree of
M.Tech CB

Biswadeep Khan (MT16124)

Will defend his thesis

Title: "Application of pattern mining on data of flavor molecules, their percepts and molecular features"

IIIT-D Faculty and Students are invited

Date: July 06th 2018 (Friday)
Time: 03.00-04.30 IST
Place: Meeting Room, 3rd Floor(NAB)

Examiner:	**Internal:**	**Debarka Sengupta**
	External/Internal:	**Debasis Dash (IGIB)**
	Advisor:	**Debajyoti Bera, Ganesh Bagler**

****

Abstract

Where does a particular food gets its flavor from? A food item is a complex combination of (flavour) molecules, in different proportions, each having their own distinctive flavors. This thesis takes a small step towards understanding the mystery of flavors by analysing the role of the functional groups present in the flavor molecules of the different ingredients present in food ingredients of different categories. We implement data mining techniques on data from FlavorDB database to arrive at novel insights on functional groups and flavor combinations. We started by modeling the underlying factors as a tripartite network among molecules, flavor terms and functional groups. Then we study a projection of this graph to a bipartite network of flavor terms and functional groups where the strength of the links indicate molecules that enforce this connectivity. The bicliques in this bipartite network revealed tighter association of flavor terms and functional groups and the corresponding molecules. We identified various co-occurring associations that are the basis of emergent flavors that characterize different categories of food. The tight associations can also be used for explaining the role of functional groups in specifying a combination of flavours. Association rule mining is a well-known data mining technique to identify causal structures in the bipartite data. We used this technique to discover strong relations among flavour terms and functional groups in four major categories of food, namely, fruit, vegetable, spice and meat. We observed that patterns discovered using both biclique analysis and association rule mining were distinct for different categories of food. Thus these patterns can be used to characterize a food molecule based on its flavour and functional group. Lastly, we undertook the task of prediction of flavour of a molecule using several molecular descriptors like number of hydrogen bond, number of rings, etc. To tackle the relatively large number of flavour terms, we first used spectral clustering to construct 15 clusters of flavour terms. Then we used four different multi-class classification methods (decision tree, K-nearest neighbour, random forest, and neural network) with 10-fold cross-validation and could obtain an accuracy between 45% and 55% of predicting flavours of molecules. While it is undeniable that flavour a food item arises due to a multitude of reasons, the techniques and observations of this thesis shows that data mining techniques can be used to explain, to some extent, how molecular composition can affect flavour of flavour molecules.