

**COLLOQUIUM SERIES:
FALL 2015**

CLUSTERING CATEGORICAL DATA VIA ENSEMBLING METHODS

Abstract:

We propose an ensemble approach to clustering categorical data. The proposed ensemble method is based on hierarchical clustering under average linkage. We give a rationale for why our procedure does well in low dimensions. This is supported by extensive computational comparisons with other methods using simulated and real data. Our method for low dimensional categorical data extends to high dimensional categorical data by using an extra level of ensembling. This minimizes the effect of the Curse of Dimensionality that tends to equalize the distances between any two points as dimension increases. A further extension of our ensembling method permits the vectors of categorical outcomes to have different dimensions.

***This presentation is part of joint with Bertrand Clarke and Jennifer Clarke from University of Nebraska-Lincoln.
<http://arxiv.org/pdf/1506.07930v1.pdf>

WWW.STAT.IOWA.EDU

**DEPARTMENT OF STATISTICS & ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
319-335-0712**

PRESENTER

Saeid Amiri
Department of Statistics and
Actuarial Science
The University of Iowa



PROFESSIONAL INTERESTS:

Clustering
Categorical data
Ensemble methods
High dimensional data

WHEN

**October 15, 2015
3:30 p.m.**

WHERE

61 Schaeffer Hall

RECEPTION

**241 Schaeffer Hall
3:00 p.m.**

