

A Bayesian Approach to Collaborative Dish Selection

Team 10

Introduction

As anyone who has ever planned a catered event can attest, attempting to satisfy the various palates, dietary requirements and tastes of a group of diners can be a daunting task. This is particularly true given the exponential number of dishes which can be created from a small number of ingredients, as well as hard constraints such as allergies and religious beliefs. Many professional catering services handle this problem by allowing guests to select from a very limited menu. We introduce a dish recommendation system based on Bayesian Networks modeling user preferences. We predict the meals from a data base of recipes that most likely match the varied tastes of the customers, using a limited set of ingredients. This type of expert system would be of great use to a catering service or restaurant which needs to rapidly decide on a small number of dishes which would be acceptable for a large dinner party, given diverse requirements and preferences.

Related Work

Boekel and Corney propose using Bayesian Networks to model consumer needs in food production chains [5] [1]. Janzen and Xiang propose an intelligent refrigerator capable of generating meal plans based on inventory and past food choices [2]. Bayesian networks have also been applied to recommendation systems before in on-line social networks [4] making predictions of the form “if you bought those items what is the probability you would like to buy that”. We suggest that these approaches are limited in that they only consider the preferences of a single (or supposed ‘typical’) user rather than a group.

Approach

The approached problem is to pick a single meal which best meets the requirements and tastes of different people dining together. We learn a predictive bayesian net from a survey distributed to participants of the meal as training data in order to capture their preferences. The dishes in the questionnaire are selected such that all ingredients are covered. The participants rate each dish on a scale from one to ten and give additional information like vegetarians. For new dishes we then predict the maximum likelihood rating given our

model. In the following we will describe our approach in detail. First we will discuss the data selection, then the modeling of the user preference and in the end how we trained the modeled net from gathered data and predicted the value for different recipes.

Data accuisition We first accumulated a diverse collection of license-free sample recipes from web sites such as *Darkstar’s Meal-Master Recipes* (<http://home.earthlink.net/~darkstar105/>). Next, we converted these recipes from flat text files to well-formed XML using the Krecipes application for Debian Linux. Finally, we created a representative data set representing several diners’ preference for 24 of these recipes, using a simple survey of the type ‘rate on a scale of 1 to 10, 10 being favorite and 1 being least favorite’. Furthermore, users were allowed to specify a vegetarian or nut-free meal preference.

Knowledge Engineering We model the diners’ various taste preferences preferences using a Bayes net. In the first layer in Figure 1 we capture control variables such as vegarian or allergy to nuts. These were modeled as boolean variables, with ‘true’ indicating the presence of a constraint.

Our model consists of 4 layers, each modeling a different aspect of taste or dietary requirement.

Layer 1 General meal preferences such as being vegetarian or being allergic to nuts.

Layer 2 The second layer models a general preference towards different food categories like vegetables or meat. As one can see, the food categories are dependent on the general meal preference. For example being vegetarian will exclude meat and will support vegetables.

Layer 3 Specific flavors and ingredients. Each ingredient is conditioned by the food category to which it belongs.

The overall net is shown in Figure 1. Given a recipe with a list of ingredients $I = i_1, \dots, i_n$ and a Bayesian network capturing user preferences we can calculate the probability of users liking the dish given the probabilities of liking each ingredient.

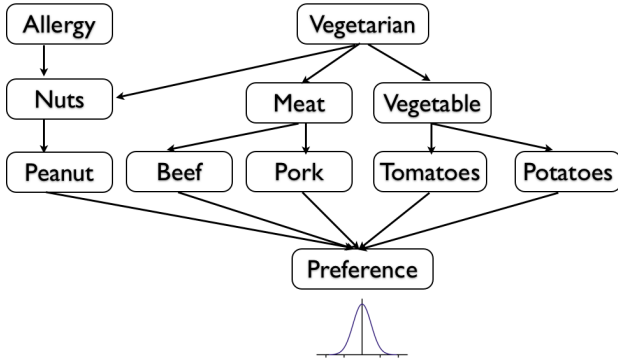


Figure 1. Our Bayesian net modeling user preferences

Learning and Predicting In order to estimate the model parameters, the system will be trained with statistics about taste and preferences given a set of dishes with ratings from multiple users. From that information we can directly calculate the probabilities for the ingredients using Maximum Likelihood Learning [3].

In order to model food preferences, we implemented a custom Bayesian net library in java with minimal use of third party libraries (e.g. for XML input). We chose to implement our own Library, for maximum flexibility and to ensure that the learning algorithm functions precisely as follows:

The library uses the sum-product algorithm for inference and maximum likelihood learning for parameter estimation. In our implementation we support discrete as well as continuous probability distributions. Discrete distributions can be modeled as tables or as trees. In our implementation only continuous distributions with discrete parents are supported. A continuous distribution is then modeled as a mapping of all possible combination of it's parents to a gaussian. Given a data set, the parameters of a discrete variable X are estimated as

$$P(X = x | Y_1 = y_1, \dots, Y_2 = y_2) = \quad (1)$$

$$\frac{N(X = x | Y_1 = y_1, \dots, Y_2 = y_2)}{N(Y_1 = y_1, \dots, Y_2 = y_2)} \quad (2)$$

where $N(A)$ is the number of times event A occurs in the data set.

Evaluation

The application model will be trained using a sparse subset (50%) of the survey data and the optimization problem solved for the inferred constraints. As shown below, the calculated preferences for recipes which were not used to train the Bayes net are quite close to the

actual survey data, which essentially reflects the following preferences (Sample ratings are on a 1-10 scale):

Diner 1 No allergies, prefers all dishes equally (5)

Diner 2 Vegetarian, meat dishes are (1), remainder are (9)

Diner 3 Nut Allergy, prefers meat (6) to vegetarian (4) to dessert (3)

Diner 4 No allergies, prefers Pork and Desserts (9), remainder are (3)

Next, we calculate the correlation between the application's ranking of all dishes and the actual ranking as determined by the user surveys. We suggest that a high degree of correlation indicates that the system has the potential to accurately appraise constrained group food preferences for dishes which are not part of the survey, given sufficiently detailed recipe information. As shown, the estimated food preferences are quite close to the actual mean ratings over all diners for the dishes which were not used to train the Bayes net. The root mean-square-error for calculated vs. surveyed meal preferences is approximately 1.0.

Southwest Smoothie:		
DAIRY	1.2	5.5
Bayou Shrimp Creole:		
TOMATO		3.75
Crab Burgers:		
EGGS		3.75
Broiled Flounder:		
GENERIC NUTS, EGGS		3.75
Baked Steak And Lima Beans:		
TOMATO, SUGAR		3.75
Eggplant Lasagna:		
GLUTEN		5.25
Salisbury Steak with Mushroom Sauce:		
GLUTEN, DAIRY, BEEF		3.75
Meatless Loaf:		
SPICE		5.25
Lemon Pork Chops:		
PORK, SUGAR		5.25
Fava Bean Burgers:		
EGGS, POTATO		5.25
Angel Hair Pesto Primavera:		
GENERIC NUTS, SPICE		5.25
Kahlua Cake		4.5

References

- [1] David Corney. Designing Food with Bayesian Belief Networks. *ACDM*, 2000.

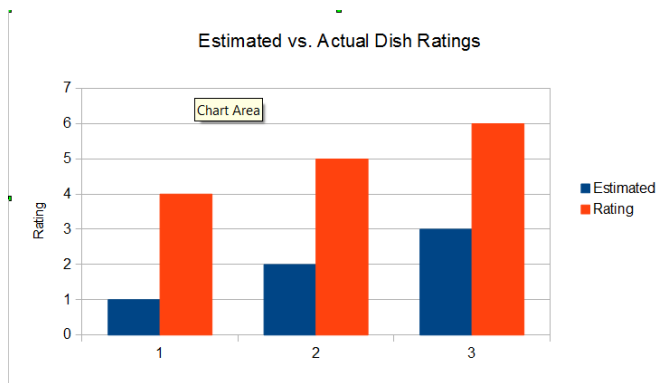


Figure 2. Estimated vs. surveyed dish ratings

- [2] M. Janzen and Y. Xiang. Probabilistic Reasoning in Meal Planning in Intelligent Fridges. *16th Conference of the Canadian Society for Computational Studies of Intelligence*, 2003.
- [3] Kevin Murphy. A Brief Introduction to Graphical Models and Bayesian Networks. <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>, 1998. [Online; accessed 22-Feb-2012].
- [4] Tran The Truyen, Dinh Q. Phung, and Svetha Venkatesh. Preference Networks: Probabilistic Models for Recommendation Systems. *Proceedings of the Sixth Australasian Conference on Data Mining*, 2007.
- [5] Stein A. van Boekel, M.A.J.S. and A.H.C. van Bruggen. Bayesian Statistics and Quality Modelling in the Agro-food Production Chain. *Proceedings of the Frontis workshop*, 2004.