

Profiling US Restaurants from Billions of Payment Card Transactions

Himel Dev
University of Illinois (UIUC)
Champaign, IL, USA
hdev3@illinois.edu

Hossein Hamooni
Facebook Inc.
Menlo Park, CA, USA
hamooni@fb.com

Abstract—A payment card (such as debit or credit) is one of the most convenient payment methods for purchasing goods and services. Hundreds of millions of card transactions take place across the globe every day, generating a massive volume of transaction data. The data render a holistic view of cardholder-merchant interactions, containing insights that can benefit various applications, such as payment fraud detection and merchant recommendation. However, utilizing these insights often requires additional information about merchants missing from the data owner’s (i.e., payment company’s) perspective. For example, payment companies do not know the exact type of product a merchant serves. Collecting merchant attributes from external sources for commercial purposes can be expensive. Motivated by this limitation, we aim to infer latent merchant attributes from transaction data. As proof of concept, we concentrate on restaurants and infer the cuisine types of restaurants from transactions. To this end, we present a framework for inferring the cuisine types of restaurants from transaction data. Our proposed framework consists of three steps. In the first step, we generate cuisine labels for a limited number of restaurants via weak supervision. In the second step, we extract a wide variety of statistical features and neural embeddings from the restaurant transactions. In the third step, we use deep neural networks (DNNs) to infer the remaining restaurants’ cuisine types. The proposed framework achieved a 76.2% accuracy in classifying the US restaurants. To the best of our knowledge, this is the first framework to infer the cuisine types of restaurants by analyzing transaction data as the *only* source.

Index Terms—payment card transaction, restaurant profiling, weak supervision, restaurant embedding

I. INTRODUCTION

Payment cards (primarily debit and credit), with billions of active users, are among the most popular payment methods across the globe. In the US alone, millions of people regularly use debit and credit cards to pay for goods and services. Payment processing companies handle these transactions and record their attributes (e.g., cardholder id, merchant name, and authorized amount), generating a massive volume of transaction data. Transaction data render a holistic view of cardholder-merchant interactions, containing insights into the behavior of cardholders and the business of merchants. The insights can benefit a variety of applications, such as payment fraud detection and merchant recommendation.

While transaction data accommodate valuable insights, utilizing these insights often require auxiliary information about

merchants that are not present in raw data. For instance, in payment fraud detection, it is critical to know merchants’ categories—the type of goods or services they provide—to classify and track purchases. Recommendation services also require characterizing merchant attributes, such as business category and sub-categories (the type of goods or services), market positioning (high-end vs. low-end), and business hours (when to visit). Broadly, incorporating merchant attributes can boost the performance of machine learning models that empower today’s recommendation and anomaly detection services.

However, acquiring merchant attributes is challenging, especially for commercial purposes. Even payment processing companies that possess the most detailed view of transactions do not have access to all required attributes. One possible way to acquire the missing attributes is to collect them from crowd-sourcing companies (e.g., Yelp, Foursquare), which generate merchant profiles from customer reviews. However, such data acquisition has several limitations. First, it is costly to acquire merchant information for commercial purposes. Second, debit and credit cards are commonly used in many countries where collecting merchant information through crowd-sourcing may not be possible. For example, Yelp currently operates in 32 countries, whereas Visa cards are accepted in more than 200 countries and territories. Third, the collected information may become outdated, as new merchants appear or old merchants go out of business. These limitations motivate us to generate merchant profiles by analyzing transactions. Unlike user reviews, transaction data do not contain a large corpus of texts or images to extract merchant attributes. Consequently, generating merchant profiles from transaction data requires inferring attributes through other means.

Present Work. The present work contributes to our broad goal of inferring latent merchant attributes from transaction data. Though raw transactions do not contain much information about merchants, the transaction data as a whole contains rich signals, which have the potential to reveal a multitude of merchant attributes. As *proof of concept*, we concentrate on restaurants and infer the cuisine types of restaurants from transactions. Restaurant is perhaps the most popular merchant category, with which most cardholders have regular transactions. However, while processing any restaurant transaction, a payment company does not know what type of cuisine the

*During this work, both authors were employed at Visa Research.

restaurant serves. This information can be useful for providing restaurant recommendations and improving fraud detection.

Restaurant Recommendation. Cuisine types play a critical role in restaurant recommendation. For example, consider a transaction-based restaurant recommendation system that provides recommendations without being aware of restaurants’ cuisine types. A standard solution to develop such a recommendation system is to employ user-based collaborative filtering, which generates recommendations based on the similarity of transactions among cardholders. While practical, this solution comes with considerable risk—imagine recommending steakhouses to vegetarians who visited many Mediterranean restaurants, as Mediterranean cuisine is popular among both vegetarians and meat-lovers. Incorporating cuisine types of restaurants into the recommendation algorithm can prevent such misreckoning. Cuisine types can also be useful for addressing the cold-start problem in restaurant recommendation.

Fraud Detection. Cuisine types of restaurants also have the potential to inform fraud detection systems. For example, suppose a fraud detection system has access to the cuisine types of restaurants. In that case, it can use this information in conjunction with transaction data to generate a cuisine signature for each cardholder. This signature captures how likely the cardholder is to visit different types of restaurants. Then, while processing a restaurant transaction, the system can examine if the restaurant aligns with the cardholder’s cuisine signature, among other signals, to detect if the transaction is a fraudulent one.

This paper presents a framework to infer the cuisine types of restaurants from billion-scale payment card transactions. Please note that the intended use of the framework includes commercial purposes. For this reason, we can not use external sources (e.g., Google, Yelp) for data acquisition. The proposed framework has three steps: 1) weakly-supervised label generation, 2) statistical feature and neural embedding extraction, and 3) deep neural network based classification. In the first step, we programmatically generate cuisine labels for restaurants via weak supervision. We first compile a cuisine taxonomy containing the major cuisine types and a list of keywords for each cuisine. We then use the keywords as common patterns to label restaurants based on their names. For example, if the keyword “Peking” appears in a restaurant’s name, we recognize it as a Chinese restaurant. Through bootstrapped label expansion, our weakly-supervised approach can generate cuisine labels for 35% of US restaurants. We infer the remaining US restaurants’ cuisine labels by employing large-scale feature engineering followed by deep neural network based classification. In the second step, we extract a wide variety of statistical features and neural embeddings for half a million US restaurants, from four billion restaurant transactions. Our statistical features encompass a wide range of business information about restaurants, such as location, business hours, pricing information, tipping culture, serving capacity, expected party size, customer visitation patterns, and customer loyalty. We also model transaction data as a bipartite graph between customers and restaurants and then

employ state-of-the-art neural language models to generate two sets of restaurant embeddings. The embeddings capture the semantic similarity between restaurants at two distinct levels: the semantic similarity between any two restaurants that an individual customer visits (micro-level), the semantic similarity between two restaurants based on their entire customer base (macro-level). We further apply neural language models on restaurant names (excluding the words used in labeling) to generate another set of restaurant embeddings. Overall, we generate three sets of restaurant embeddings to represent the latent characteristics of restaurants. In the third step, we design several deep neural networks (DNNs) to predict remaining restaurants’ cuisine types based on the statistical features and embeddings generated in the second step. Our best performing network is a deep feedforward network with residual connections. The network achieved a 76.2% overall accuracy in classifying US restaurants, where per class accuracy is high for all classes, including minority cuisine types such as Mediterranean cuisine. To the best of our knowledge, this is the first attempt to profile restaurants in the mostly untapped space of payment card transactions.

II. PROBLEM SETTING

This work aims to develop a framework for inferring the cuisine types of restaurants from debit and credit card transactions. Cuisine type information is essential to payment processing companies for providing user services, notably, restaurant recommendation and payment fraud detection.

Feature Engineering. A secondary goal of this work is to extract general-purpose features that can benefit a variety of tasks beyond cuisine-based restaurant classification, such as business profiling and customer segmentation. Preferably, we want to extract interpretable features that can provide insights into the restaurant business.

Data Restriction. The intended use of this framework includes commercial purposes. Therefore, we can not use external sources (e.g., Google, Yelp) for label generation, feature engineering, and framework validation. For example, without explicit consent, Yelp API and content usage are restricted to non-commercial use [1]. Acquiring merchant information for commercial purposes can be costly.

In remaining sections we present a framework that meets the requirements above and achieves a high accuracy in inferring the cuisine types of US restaurants.

III. TRANSACTION DATASET

Our dataset contains four billion debit and credit card transactions in more than half a million US restaurants within three months. This dataset is a large subset of in-person restaurant transactions in the US (customer made payment at a restaurant using a VISA card). Each transaction in our dataset is an interaction between a cardholder and a merchant. Therefore, all attributes of each transaction come from the cardholder, the merchant, or their interaction. Specifically, our dataset contains the following information for each transaction.

- (a) **Merchant Name:** the legal name of the merchant.

- (b) **Merchant Location:** the ZIP code of the merchant.
- (c) **Timestamp:** the exact date and time of the transaction.
- (d) **Cardholder Id:** an anonymous id to group all transactions of the cardholder.
- (e) **Authorized Amount:** the amount (in USD) that the merchant initially charges.
- (f) **Settlement Amount:** the amount (in USD) that finally gets posted on the cardholder’s bank statement. For restaurant transactions, the difference between the authorized amount and the settlement amount is the tip.

IV. CUISINE INFERENCE FRAMEWORK

In this section, we describe our proposed framework for inferring the cuisine types of restaurants from debit and credit card transactions. Figure 1 presents an overview of the framework. The framework has three steps: 1) weakly-supervised label generation, 2) statistical feature and neural embedding extraction, 3) deep neural network based classification. We perform these three steps in a *closed-loop* to classify the restaurants in our transaction dataset.

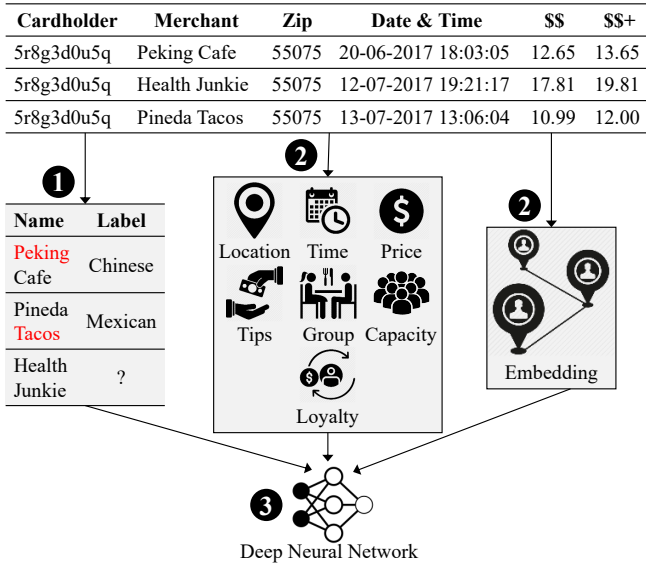


Fig. 1. Our proposed framework for inferring the cuisine types of restaurants. The framework has three steps: 1) weakly-supervised label generation, 2) statistical feature and neural embedding extraction, 3) deep neural network based classification.

A. Weakly-Supervised Label Generation

In the first step, we programmatically generate cuisine labels for restaurants via weak supervision. We compile a cuisine taxonomy along with keywords, extract new keywords from restaurant names via bootstrapping, and employ topic modeling to augment the keyword-based approach.

Cuisine Taxonomy Compilation. We compile a cuisine taxonomy based on the literature on food [2], [3] and the US restaurant industry [4]–[6]. Table I presents our cuisine taxonomy. Our taxonomy contains the ten most popular cuisine types in the US: Latin American, European, Middle Eastern & Mediterranean & African (MMA), South Asian, South East

Asian, East Asian, Grill & Steak, Fastfood, Bar, Dessert. Each of these major cuisine types (e.g., European) accommodates more fine-grained types (e.g., French). Though our taxonomy does not contain all cuisine types (e.g., Circassian), it has a high coverage for US restaurants. We compile a set of *seed* words for each cuisine type in our taxonomy. We use these words as common patterns to generate cuisine labels for restaurant names. For example, if the keyword “Peking” appears in a restaurant’s name, we recognize it as a Chinese restaurant. Currently, we have a list of 225 seed words that represent the ten major cuisine types.

TABLE I
MANUALLY CRAFTED CUISINE TAXONOMY FOR RESTAURANT CLASSIFICATION.

Cuisine Type	Subcategories
Latin American	Mexican, Cuban, Brazilian, Colombian
European	French, Italian, German, Polish, Irish
Mediterranean	Greek, Turkish
Middle Eastern	Saudi Arabian, Lebanese, Persian, Afghan
African	Moroccan, Ethiopian, Eritrean
South Asian	Indian, Pakistani, Nepalese, Bangladeshi
South East Asian	Thai, Vietnamese, Indonesian, Malaysian
East Asian	Chinese, Japanese, Korean, Mongolian
Grill and Steak	Grill, Steakhouse
Fastfood	Sandwich, Burger, Pizza
Bar	Bar, Pub, Tavern, Inn
Dessert	Ice Cream, Cafe, Bakery, Juice

Bootstrapped Label Expansion. To better utilize the cuisine information obtainable from restaurant names, we apply pattern-based bootstrapping [7]. The key idea of our approach is to extract new words (beyond seed words) from restaurant names, which can serve as highly accurate patterns to increase the coverage of labeled restaurants. We first determine the cuisine label distribution for each new word based on the labeled restaurants. We then select the words that meet three *minimum* criteria—i) *frequency*: the word needs to appear in θ_f fraction of restaurant names; ii) *precision*: if we use the word and its majority label as a labeling rule, the rule needs to be true for θ_p fraction of labeled restaurants; iii) *significance*: the ratio of labeled and unlabeled restaurants that contain the word should be θ_s . We identify 198 words that satisfy the criteria mentioned above and use these bootstrapped words as secondary patterns to label more restaurants. Altogether, using the seed and bootstrapped keywords, we label 35% of restaurants in our dataset.

Topic Modeling for Weak Supervision. To augment the keyword-based approach of weak supervision, we apply a variety of topic models on restaurant names, treating each restaurant name as a document. Conventional topic models such as LDA [8] do not work well with restaurant names,

TABLE II
THE DESCRIPTION OF STATISTICAL FEATURES AND NEURAL EMBEDDINGS EXTRACTED FROM RESTAURANT TRANSACTIONS.

Feature Type	Description
Pricing	The deciles of authorized amount in transactions
Tipping culture	The deciles of (settlement amount - authorized amount) in transactions
Serving capacity	The deciles of hourly transaction count
Party size	The proportion of transactions for different party size
Party pricing	The average authorized amount for different party size
Temporal pattern I	The distribution of number of transactions over days of the week
Temporal pattern II	The distribution of number of transactions over the hours of weekdays
Temporal pattern III	The distribution of number of transactions over the hours of weekends
Customer revisitation	The deciles of the number of revisits by the customers
Customer loyalty	The deciles of the number of restaurants visited by the customers
Location	The digits of restaurant zipcode and corresponding location granularity
Micro embedding	The restaurant embedding generated by applying word2vec on user documents
Macro embedding	The restaurant embedding generated by applying doc2vec on restaurant documents
Name embedding	The embedding based on restaurant name

resulting in incoherent topics, as measured via UMass coherence score [9]. We observe three critical challenges in applying topic models— i) *monolith*: many restaurant names consist of a single word; ii) *sparsity*: sparse word co-occurrence patterns in restaurant names; iii) *long-tail*: long-tail distribution of words appearing in restaurant names. To address the monolith issue, we adopt the concept of *sprinkling* [10]—adding artificial words (known cuisine types) to boost the association in the topic model. To address the sparsity issue, we adopt the bi-term topic model (BTM) [11] that learns topics by directly modeling the generation of word co-occurrence patterns known as bitersms. BTM has two primary advantages over conventional topic models: i) BTM explicitly models word co-occurrence patterns, instead of word documents, which improves the topic learning process; ii) BTM uses the aggregated patterns in the whole corpus for learning topics, which solves the problem of sparse patterns at document-level. To address the long-tail issue, we perform *stratified sampling* on bitersms before passing them to BTM. Using BTM as our topic model, we vary the topic model parameter k (the number of topics) from 5 to 50, while assigning the value $\frac{50}{k}$ to parameter α (document-topic density) and 0.1 to parameter β (topic-word density), as suggested by Griffiths et al. [12]. We then compute the UMass coherence score [9] for each set of topics. The resultant topics (cuisine classes) are coherent (high UMass score) and consistent with the cuisine classes generated by our keyword-based approach.

B. Statistical Feature and Embedding Extraction

In the second step, we perform extensive feature engineering for large-scale transaction data. Table II presents the description of the features that we extract from restaurant transactions. Broadly, our features fall into two categories: statistical features that capture business information about restaurants, and embeddings that capture restaurant’s latent characteristics based on customer-restaurant interactions.

Statistical Feature Extraction. We extract a wide variety of statistical features from transactions based on the literature on the restaurant industry [13]–[25]. These features encompass a wide range of business information about restaurants, such as location, business hours, pricing information, tipping culture, serving capacity, expected party size, customer visitation patterns, and customer loyalty. The features provide insights into the restaurant business and benefit various tasks beyond restaurant classification, such as business profiling and customer segmentation. A key idea of our statistical feature engineering approach is to use distributions as features, enabling the framework to differentiate among cuisine types in terms of tail behavior.

Pricing: Price is one of the crucial features for characterizing a restaurant, which bears implications for cuisine type [23]. Roy [24] reported the price inequality among different cuisine types for restaurants located in New York City (NYC), e.g., people in NYC pay more for European cuisines compared to East Asian and South Asian cuisines. To capture the idiosyncrasies of pricing, we extract the deciles of the authorized amount for each restaurant. The 2nd column in Table III shows how the median price of food varies across cuisine types, for all US restaurants (median of median). The median price is: a) high for European cuisines; b) medium for Bar, Grill & Steak, and Asian cuisines; c) low for Dessert, Fastfood, Latin American, and MMA cuisines.

Tipping Culture: Tipping norms vary across restaurant types, e.g., fine dining restaurants have different norms compared to fast food restaurants. Chen et al. [25] examined how the bias towards different types of cuisines affect gratuities. The study confirms the effect of cuisine types on tipping behavior. To capture customers’ tipping behavior for a restaurant, we extract the deciles of the difference between the settlement amount and the authorized amount. The 3rd column in Table III shows how the median of tips varies across cuisine types. The median tip is: a) high for Bar and European cuisines; b) medium for Grill & Steak, South Asian, and South East

TABLE III
DIFFERENCE AMONG CUISINE TYPES IN TERMS OF STATISTICAL FEATURES. PERCENTILE BASED COLOR CODING: LOW, MEDIUM, HIGH

1. Cuisine Type	2. Price (Median)	3. Tips (Median)	4. Serving Capacity (Mean)	5. Customer Revisit (90th Percentile)	6. Restaurant Tally (Median)	7. Expense/Person (Mean)	8. Single Diner (Percentage)	9. Weekend Tranx. (Percentage)
Dessert	8.13	0	24	32	19	5	46	48
European	28.89	4	23	17	17	17	45	53
Fastfood	16.45	0	19	19	17	9	42	50
Bar	22.53	4	27	30	18	13	45	54
Grill & Steak	21.91	3	26	24	18	12	44	51
Latin American	16.69	0	23	22	18	10	43	48
MMA	16.57	0	20	18	19	11	46	46
South Asian	23.87	2	18	16	18	13	40	52
South East Asian	21.88	2	17	16	19	11	35	47
East Asian	19.40	0	17	15	17	10	39	48

Asian cuisines; c) low for all other cuisines.

Serving Capacity: Serving capacity is an important variable for studying a restaurant business [15], e.g., fast food restaurants typically have higher serving capacity compared to the fine dining restaurants [21]. To capture the serving capacity of a restaurant, we extract the deciles of the number of transactions per hour. The 4th column in Table III shows how the average number of servings varies across cuisine types, for all US restaurants (median of average). Average serving capacity (number of customers per hour) is: a) high for Bar, and Grill & Steak; b) medium for Dessert, European, and Latin American cuisines; c) low for all other cuisines.

Party Size: Party size composition is an insightful feature for studying the customers of a restaurant [17], [18]. Qu et al. [14] showed that at Chinese restaurants in Indiana, the majority (57%) of customers dined with a party size of two. Estimating party size from transaction data is challenging. A bill for a dining party could be paid by one of the members or shared among the members. If one of the members pays the entire bill, there would be no explicit information about the party size. In contrast, if the members share the bill, there would be no association between their transactions. We overcome the first challenge by estimating party size from the authorized amount. We can not address the second challenge due to the weak association between transactions. To estimate party size, we apply a separate Gaussian mixture model (GMM) on each restaurant’s authorized amounts. Our Gaussian mixture model is a weighted sum of K component Gaussian densities as given by the equation,

$$p(x|\theta) = \sum_{i=1}^K \phi_i \mathcal{N}(\mu_i, \sigma_i),$$

where x is the vector of authorized amounts, ϕ_i is the

mixture weight (prior probability) of i -th component, and $\mathcal{N}(\mu_i, \sigma_i)$ is the Gaussian density of i -th component. We determine the *optimal* number of mixture components for GMM by using the Akaike information criterion (AIC). We use the proportion of transactions for different party sizes, and the corresponding average authorized amounts as party composition features. The 7th column in Table III shows how the average expenditure per person varies across cuisine types. Consistent with pricing, the average expenditure per person is highest for European cuisine and lowest for dessert. The 8th column in Table III shows how the percentage of single diner varies across cuisine types. Notice that the single diner percentage is lower for the Asian cuisines compared to other cuisine types. This finding is consistent with the norm that Asian restaurants are popular for group eating.

Temporal Patterns: Temporal patterns such as business hours and busy times are critical indicators for understanding a restaurant business [13], e.g., bars typically operate during night time, whereas brunch spots operate during day time. To capture the temporal patterns of a restaurant’s operation, we extract the distribution of transactions over days-of-week and hours-of-day. We further divide the hours-of-day into hours-of-weekday and hours-of-weekend. The 9th column in Table III shows how the percentage of transactions on weekends varies across cuisine types, for all US restaurants. Notice that Bars and European restaurants attract more of their customers during weekends.

Customer Revisitation: Researchers have acknowledged the importance of understanding the differences between first-time customers and repeat customers in restaurants [19]. Barber et al. [20] showed a strong correlation between the food type and repeat visitation of customers. To capture the customer revisitation tendency in a restaurant, we extract the deciles of the number of revisits (repeat transactions) by the

customers. The 5th column in Table III shows how the number of revisits for top customers (on and above 90th percentile) varies across cuisine types. Notice that frequent customers have a strong revisit tendency to bars and dessert places.

Customer Loyalty: Customer loyalty can be useful for studying a restaurant business [16]. This feature is especially helpful for understanding the customers and their endemicity. To capture the loyalty of customers of a restaurant, we extract the deciles of the number of restaurants visited by the customers. The 6th column in Table III shows how the median number of restaurants visited by customers varies across cuisine types. Notice that there is not much difference among the cuisine types in terms of customer base loyalty.

Location: Location has been a critical factor for characterizing a restaurant [22]. As a restaurant business is often established based on potential customers' preference in the locality, location plays a crucial role in restaurant profiling. We extract the digits of restaurant zip code and corresponding location granularity to capture a restaurant's locality.

Restaurant Embedding Generation. Restaurants in our dataset can be represented in a lower-dimensional space based on the inherent structure imposed by their characteristics, such as cuisine type, price range, and location. Cardholders can also be represented in a lower-dimensional space based on their traits, such as cuisine affinity, price sensitivity, and location. The transactions in our dataset are interactions between samples from these two spaces, where certain interactions are plausible while others are unlikely to happen. To capture the interactions between cardholders and restaurants, we model transaction data as a bipartite graph. We then propose two hypotheses that capture the plausibility of the interactions in this graph at the micro and macro level.

Hypothesis I: *The plausibility of interaction between a customer and a restaurant is a function of the customer's preferences and the restaurant's attributes.*

Hypothesis II: *The plausibility of interaction between a group of customers and a restaurant is a function of the collective taste and the restaurant's attributes.*

The subtle distinction between the two hypotheses lies in their application: the first hypothesis applies to individuals (micro), and the second hypothesis applies to groups (macro). The first hypothesis implies that the compatibility between a customer's preferences and a restaurant's attributes is a good predictor of whether the customer will visit the restaurant. For example, a vegetarian is likely to visit an Indian restaurant. The second hypothesis implies that the type of customers who visit a given restaurant is a good predictor of its attributes. For example, a restaurant is unlikely to be a steakhouse if many of its customers are vegetarian.

Based on the hypotheses mentioned above, we aim to represent the restaurants in our dataset in a low-dimensional space by applying state-of-the-art representation learning techniques.

There is a variety of methods that can be used for learning lower-dimensional representations—notably neural language models, such as word2vec [26] and its extension doc2vec [27]. These models have been successfully adopted in a variety of domains, including product representation in eCommerce [28]. In this work, we utilize word2vec and doc2vec to create two distinct sets of restaurant embeddings based on our micro and macro hypotheses. Details about the parameters of the micro (word2vec) and macro (doc2vec) model can be found in the experimental evaluation (impact of hyperparameters).

Micro Embedding: We represent each customer as a document and each restaurant as a word. Accordingly, each transaction by a customer corresponds to a word (restaurant) in a document (the customer's ledger of restaurants). The words appear chronologically in the documents, as per the day and time of transactions. After creating the documents, we apply word2vec on these documents. As per our representation, words refer to restaurants, and consequently, the word embeddings generated by word2vec are restaurant embeddings. These embeddings capture our micro hypothesis: the context of a restaurant in a customer's ledger is defined by other restaurants visited by the customer within the same time window.

Macro Embedding: We represent each restaurant as a document and each customer as a word. Accordingly, each transaction in a restaurant corresponds to a word (customer) appearing in a document (the restaurant's ledger of customers). These words appear chronologically, as per the day and time of transactions. After creating the documents, we apply doc2vec on these documents. As per our representation, documents refer to restaurants, and consequently, the document embeddings returned by doc2vec are restaurant embeddings. These embeddings capture our macro hypothesis: the context of a customer in a restaurant is defined by other customers visiting the restaurant within the same time window.

Name Embedding: Apart from the two sets of embeddings learned from the customer-restaurant interactions; we generate another set of embeddings from restaurant names. Name embedding has been successfully utilized in other domains, such as social media [29]. In this paper, we generate name embeddings to utilize the remaining (non-labeling) words in restaurant names. To this end, we first remove the labeling words (used in step 1) from restaurant names. We then use pre-trained word embeddings (GloVe [30]) to generate name embeddings for restaurants. Specifically, to generate the name embedding of a restaurant, we combine the embeddings corresponding to its words via max pooling.

C. DNN Based Classification

In the third step, we classify the unlabeled restaurants in our dataset. To this end, we create several deep neural networks (DNNs) to complement our extracted features. Some of these networks mimic prominent DNN models deployed in large-scale industry setup, such as wide and deep [31], deep and cross [32]. Here we only discuss the architecture of our best-performing DNN model: a deep feedforward network with

residual connections. Figure 2 presents our best performing DNN model. The best performing DNN consists of several parallel paths, each corresponding to a different feature. First, we pass each of our features (vectors of real-numbers) through two hidden layers, where the layers have residual connections [33]. We then concatenate the resultant layers to create a single concatenated layer. Finally, we pass the concatenated layer through two more hidden layers and apply SoftMax activation at the output layer.

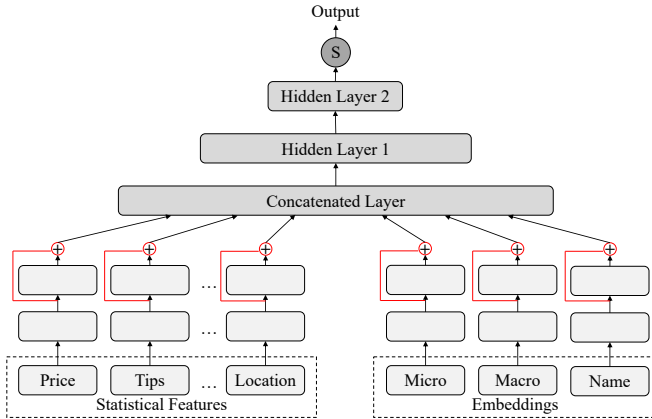


Fig. 2. Architecture of the best performing DNN model: a deep feedforward network with residual connections (marked in red).

V. EXPERIMENTAL EVALUATION

In this section, we provide a comprehensive evaluation of our cuisine inference framework. We conduct extensive experiments on a large-scale transaction dataset to validate the effectiveness of the framework. Specifically, we focus on answering the following research questions.

- 1) How effective is our framework in inferring the cuisine types of restaurants?
- 2) How useful are different feature types for classifying restaurants?
- 3) What are the impacts of different hyper-parameters in overall performance?

A. Experimental Setup

We first describe our experimental setup: evaluation dataset and DNN models.

Evaluation Dataset. Our transaction dataset contains $\sim 4,000,000,000$ transactions in $\sim 650,000$ restaurants. A sizable fraction of these restaurants are chains, e.g., McDonald’s, Subway, and KFC. For the purpose of evaluation, we exclude the chain restaurants. The final evaluation dataset contains $\sim 500,000$ restaurants. We generate cuisine labels for 35% of these restaurants using our weakly-supervised method. We then split the labeled restaurants into training (80%) and testing (20%) sets. We used 5-fold cross-validation for tuning the hyperparameters. In addition to the weakly supervised restaurants, we manually label another 1000 restaurants to create a holdout set.

DNN Models. To demonstrate the effectiveness of our framework, we develop several DNN models as follows.

- (a) **Wide & Deep:** This is a variant of the wide and deep network [31] that captures the interaction between features.
- (b) **Deep & Cross:** This is a variant of the deep and cross network [32] that applies feature crossing.
- (c) **Shallow Feedforward:** This is a feedforward neural network with two hidden layers. We first pass each feature through a hidden layer, concatenate them, and finally pass them through another hidden layer.
- (d) **Deep Feedforward:** This is a feedforward neural network with four hidden layers. We first pass each feature through two hidden layers, then concatenate them, and finally pass them through two more hidden layers.
- (e) **Deep Feedforward with Residual:** This is a deep feedforward network with residual connections [33]. We first pass each feature through two hidden layers (with residual connection between the two layers), concatenate them, and finally pass them through two more hidden layers.

B. Effectiveness of Our Framework (RQ1)

We now discuss the effectiveness of our framework in inferring the cuisine types of restaurants.

DNN Performance Comparison. We compare the performance of our DNN models for inferring the cuisine types of restaurants. Table IV reports the accuracy of our DNN models. These scores are based on the best configuration of hyperparameters (dropout, batch size, and learning rate) determined through a grid search. From the results, we make the following observations.

- 1) The Deep Feedforward network outperforms the Shallow Feedforward network as expected. The performance gap between the two is significant.
- 2) Surprisingly, the Deep Feedforward network also outperforms the Wide & Deep and Deep & Cross networks. In fact, the Wide & Deep network performs worse even compared to the Shallow Feedforward network.
- 3) Adding residual connection between layers improves the performance of DNNs, as demonstrated by the Deep Feedforward with Residual network.
- 4) While the performance gain from using the best-performing architecture may seem small in a relative scale, the gain is significant in absolute terms: 1% improvement implies correctly classifying another 5000 restaurants.

Pre-trained Embedding Baseline. VISA previously generated a set of general-purpose merchant embeddings for all merchants (including restaurants). We used the pre-trained restaurant embeddings in conjunction with a deep feedforward network to develop a baseline for inferring the cuisine types of restaurants. This baseline achieved an accuracy of 0.51, implying that the proposed framework provides a 49% improvement over the baseline.

Non-Neural Baselines. For completeness, we also compare the performance of our DNN models with several non-neural

TABLE IV
COMPARISON OF ACCURACY AMONG DNN MODELS.

Neural Architecture	Accuracy
Wide & Deep	0.740
Deep & Cross	0.746
Shallow Feedforward	0.743
Deep Feedforward	0.756
Deep Feedforward with Residual	0.762

baselines. Our non-neural baselines include logistic regression and SVM with different kernel functions (linear, polynomial, rbf, and sigmoid). The best performing non-neural baseline is the multi-class SVM with a linear kernel, which achieves an accuracy of 0.351.

Top-k Accuracy. We investigate if further improvement in accuracy can be achieved by increasing the number of predictions. We compute top-k accuracy for our best-performing model, where correct prediction is assumed if the ground truth label appears in the top-k predictions. The top-2 and top-3 accuracy scores for our best-performing model are 0.879 and 0.921, respectively.

Confusion Matrix. As our dataset is highly imbalanced (class imbalance ratio up to 16:1), we study per class accuracy to better understand performance. Figure 3 presents the normalized confusion matrix for the best performing model in the validation test. Notice that our framework leads to a *balanced* confusion matrix, where per class accuracy is high for all classes, including minority classes such as South Asian.

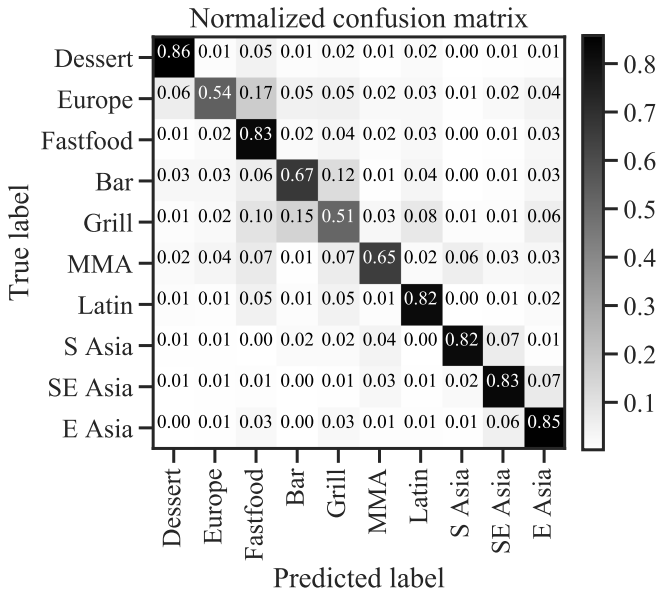


Fig. 3. Confusion matrix for the best performing model in the validation test. Notice that our framework leads to a *balanced* confusion matrix, where per class accuracy is high for all classes.

Holdout Performance. As we use the weakly-supervised labels as ground truth, the labels may introduce some bias.

To investigate if such bias affects our final results, we use the manually labeled holdout set to review our framework’s performance. We found that our framework’s performance in the holdout set is comparable with the performance in the validation set.

Model Stability. As transaction data is time-varying, we investigate if our framework’s predictions remain stable over time. To this end, we use another evaluation dataset (with six billion transactions) based on a different three months period. We generate predictions for unlabeled restaurants based on both datasets and compare the results. We observe that the predictions of our model remain stable over time.

C. Importance of Features (RQ2)

We perform an ablation study to understand the importance of different features in cuisine-based restaurant classification. Table V reports the change in accuracy scores obtained by removing one feature at a time. We observe that the neural restaurant embeddings are the most useful features for cuisine type prediction. This finding implies that the embeddings successfully capture the latent characteristics of restaurants. Among the embeddings, the contribution of name embedding is particularly noteworthy. It shows that the remaining words (that have not been used for labeling) in restaurant names contain strong signals about cuisine type. The micro embedding that captures compatibility between a customer and a restaurant is another strong indicator, implying that many customers have a strong preference towards certain cuisine types. The macro embedding is also useful in cuisine type prediction, confirming the taste similarity among the customers who visit a given restaurant. Among the statistical features, temporal patterns play a crucial role, indicating that cuisine types impact the business hours and busy times. Other prominent statistical features are price, location, and customer visitation patterns. As evidenced by the decline in accuracy score, all of the proposed statistical features are also useful in cuisine type prediction.

TABLE V
ABLATION STUDY OF FEATURES: CHANGE IN ACCURACY AFTER REMOVING EACH FEATURE.

Removed Feature	Accuracy (Change)
Name embedding	0.689 (-0.073)
Micro embedding	0.720 (-0.042)
Macro embedding	0.750 (-0.012)
Transactions over hours-of-weekend	0.750 (-0.012)
Transactions over hours-of-weekday	0.751 (-0.011)
Transactions over days-of-week	0.751 (-0.011)
Price quantiles	0.753 (-0.009)
Zipcode based location hierarchy	0.754 (-0.008)
Tips quantiles	0.755 (-0.007)
Revisitation quantiles	0.755 (-0.007)
Average spending for party size	0.756 (-0.006)
Customer count quantiles	0.757 (-0.005)
Affinity quantiles	0.757 (-0.005)
Party size composition	0.758 (-0.004)

D. Impact of Hyperparameters (RQ3)

There are several sets of hyperparameters in our framework that may impact the overall performance, notably, the hyperparameters of the DNN model and the hyperparameters associated with the neural embeddings. To determine the impact of these hyperparameters, we perform a grid search over the space of hyperparameters. Table VI shows the optimal configuration of hyperparameters (in bold). We found that two hyperparameters have a profound impact on performance: batch size (smaller batch size leads to significant improvement in performance) and the dimension of micro embedding (the dimension size significantly affects its contribution).

TABLE VI
OPTIMAL CONFIGURATION OF HYPERPARAMETERS (IN BOLD).

DNN	
Dropout	0.0 , 0.1, 0.2, 0.3
Batch Size	128 , 256, 512, 1024
Learning Rate	0.001, 0.01, 0.1 , 1.0
Micro Embedding	
Dimension	50, 100, 200 , 300
Window Size	5, 10, 20 , 30
Negative Sampling	5, 10, 20 , 30
Macro Embedding	
Dimension	50, 100 , 200, 300
Window Size	50, 100, 200 , 300
Negative Sampling	5, 10, 20 , 30

VI. DISCUSSION

We make the following observations from our results: 1) names are informative in identifying the cuisine types of restaurants; 2) different cuisine types considerably differ across the business dimensions (such as pricing, tipping culture, and serving capacity); 3) it is difficult to distinguish certain cuisine types (say South Asian and South East Asian) from one another. Now, we qualitatively analyze these results and present the implications of our findings.

Cuisine Information in Restaurant Names. Restaurant names often contain rich information about the underlying cuisine types. Recall that, with just 423 keywords (225 seed, 198 bootstrapped), we could label 35% of US restaurants. Further, we validated the efficacy of the labeling through a holdout set. Altogether, the keywords provide high support and confidence in labeling restaurants. We also generated name embeddings from the remaining words in the restaurant names, which became the most informative feature in inferring the cuisine types. Overall, names have the potential to inform the latent characteristics of restaurants.

Cuisine Types across Business Dimensions. Our study reveals several insights about the US restaurant industry. Table III shows some of the insights derived from our statistical features. A careful inspection of these statistics presented reveals the inequalities across different ethnic cuisine types

in the US. For example, across all US restaurants, the median tips for European cuisine are much higher compared to the Asian cuisines. These insights could open up new threads of research on the restaurant industry.

Similarity of Cuisines. We analyze the normalized confusion matrix to get an understanding of which classes are harder to separate. We observe that it’s hard to separate the following classes: i) MMA from Fast food; ii) MMA from Grill & Steakhouse; iii) South Asian from South East Asian; iv) South East Asian from East Asian; v) European from Fast food; vi) Bar from Grill & Steakhouse; vii) Grill & Steakhouse from Fast food; viii) Grill & Steakhouse from Bar. Note that iii) and iv) are examples of kissing cuisines [2]: cuisines that are similar in terms of ingredients and flavors. In remaining cases, there are many restaurants that serve both types of cuisines. For example, many Mediterranean places serve gyros (a popular fast food), bars often include grills.

VII. RELATED WORK

Our work draws from and improves upon, prior works on restaurant recommendation.

Restaurant Profiling and Recommendation. There is a recent body of work on restaurant profiling and recommendation [34]–[38]. These works have typically relied on user ratings and reviews for generating restaurant profiles and subsequent recommendations. Bakhshi et al. [37] studied the effects of restaurant attributes, local demographics, and local weather conditions at the date of visit on online customer reviews. Zhang et al. [36] designed a unified Collective Implicit Explicit Preference Model (CIEPM) to combine the implicit (check-in) and explicit (review) preference of customers for restaurant recommendation. Lian et al. [34] used restaurant attributes, check-in data, and review text to perform a survival analysis of restaurants in China. We explore the problem of restaurant profiling in an untapped space—from debit and credit card transactions. Unlike user reviews, transaction data does not contain a large corpus of texts or images to infer attributes. Accordingly, we perform an extensive feature engineering for transaction data to infer restaurant attributes.

Studies on Restaurant Industry. Our work draws from the rich literature on the restaurant industry. In addition to the works already mentioned, our study is motivated by the seminal work of Susan Auty [39], and its followups [40]. The key finding of Auty’s work is that food (cuisine) type is a primary factor in restaurant selection by consumers. Our proposed work builds upon this premise, where we identify cuisine types of US restaurants to provide meaningful restaurant recommendations

VIII. CONCLUSION

In this paper, we present a framework for inferring the cuisine types of restaurants from debit and credit card transactions, drawing from the literature on the restaurant industry, and applied machine learning. Our framework involves three steps—programmatically generating cuisine labels for restaurants, extracting statistical features and neural embeddings

from transactions, and finally, deep neural network based classification. Our framework achieved a 76.2% accuracy in classifying the US restaurants. Our findings have implications for feature engineering for large-scale transaction data, inferring latent merchant attributes, and studying the US restaurant industry. To the best of our knowledge, this is the first framework that performs cuisine inference by analyzing transaction data as the *only* source.

ACKNOWLEDGMENT

The authors would like to thank Mahashweta Das, Konik Kothari, Manoj Reddy, Aravind Sankar, and Hao Yang for their valuable feedback in this project.

REFERENCES

- [1] "Yelp api terms of use," https://www.yelp.com/developers/api_terms, accessed: 2020-09-04.
- [2] S. Sajadmanesh, S. Jafarzadeh, S. A. Ossia, H. R. Rabiee, H. Haddadi, Y. Mejova, M. Musolesi, E. D. Cristofaro, and G. Stringhini, "Kissing cuisines: Exploring worldwide culinary habits on the web," in *Proceedings of the 26th International Conference on World Wide Web Companion (WWW Companion)*, 2017.
- [3] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási, "Flavor network and the principles of food pairing," *Scientific Reports*, 2011.
- [4] J. H. Lee, J. Hwang, and A. Mustapha, "Popular ethnic foods in the united states: a historical and safety perspective," *Comprehensive Reviews in Food Science and Food Safety*, 2014.
- [5] K. Park, "Ethnic foodscapes: Foreign cuisines in the united states," *Food, Culture & Society*, 2017.
- [6] M. Somashekhar, "Ethnic economies in the age of retail chains: comparing the presence of chain-affiliated and independently owned ethnic restaurants in ethnic neighbourhoods," *Journal of Ethnic and Migration Studies*, 2018.
- [7] S. Gupta and C. Manning, "Improved pattern learning for bootstrapped entity extraction," in *Proceedings of the Eighteenth Conference on Computational Natural Language Learning (CoNLL)*, 2014.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research (JMLR)*, 2003.
- [9] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.
- [10] S. Chakraborti, R. Mukras, R. Lothian, N. Wiratunga, S. N. Watt, and D. J. Harper, "Supervised latent semantic indexing using adaptive sprinkling," in *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.
- [11] X. Yan, J. Guo, Y. Lan, and X. Cheng, "A biterm topic model for short texts," in *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, 2013.
- [12] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences (PNAS)*, 2004.
- [13] G. A. Fine, "Organizational time: Temporal demands and the experience of work in restaurant kitchens," *Social Forces*, 1990.
- [14] H. Qu, "Determinant factors and choice intention for chinese restaurant dining: a multivariate approach," *Journal of Restaurant & Foodservice Marketing*, 1997.
- [15] C. C. Muller, "A simple measure of restaurant efficiency," *Cornell Hotel and Restaurant Administration Quarterly*, 1999.
- [16] M. A. Clark and R. C. Wood, "Consumer loyalty in the restaurant industry: a preliminary exploration of the issues," *British Food Journal*, 1999.
- [17] S. E. Kimes and S. K. Robson, "The impact of restaurant table characteristics on meal duration and spending," *Cornell Hotel and Restaurant Administration Quarterly*, 2004.
- [18] S. E. Kimes and G. M. Thompson, "Restaurant revenue management at chevys: Determining the best table mix," *Decision Sciences*, 2004.
- [19] K. Ryu and H. Han, "New or repeat customers: How does physical environment influence their restaurant experience?" *International Journal of Hospitality Management*, 2011.
- [20] N. Barber, R. J. Goodman, and B. K. Goh, "Restaurant consumers repeat patronage: A service quality concern," *International Journal of Hospitality Management*, 2011.
- [21] E. Schlosser, *Fast food nation: The dark side of the all-American meal*. Houghton Mifflin Harcourt, 2012.
- [22] J. She, A. Vassilovski, and A. Hon, "What cuisine do you like?: Improving dining preference prediction through physical social locations," in *2012 IEEE International Conference on Green Computing and Communications (GreenCom)*, 2012.
- [23] E. S. Yim, S. Lee, and W. G. Kim, "Determinants of a restaurant average meal price: An application of the hedonic pricing model," *International Journal of Hospitality Management*, 2014.
- [24] K. Ray, *The Ethnic Restaurateur*. Bloomsbury Publishing, 2016. [Online]. Available: <https://books.google.com/books?id=orHXCgAAQBAJ>
- [25] C. Chen and Y. Chen, "The impacts of different types of cuisines and restaurants on gratuities," *Journal of Revenue and Pricing Management*, 2017.
- [26] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems (NIPS)*, 2013.
- [27] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International Conference on Machine Learning (ICML)*, 2014.
- [28] M. Grbovic, V. Radosavljevic, N. Djuric, N. Bhamidipati, J. Savla, V. Bhagwan, and D. Sharp, "E-commerce in your inbox: Product recommendations at scale," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2015.
- [29] J. Ye and S. Skiena, "The secret lives of names? name embeddings from social media," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2019.
- [30] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014.
- [31] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, and H. Shah, "Wide & deep learning for recommender systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (DLRS)*, 2016.
- [32] R. Wang, B. Fu, G. Fu, and M. Wang, "Deep & cross network for ad click predictions," in *Proceedings of the ADKDD*, 2017.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [34] J. Lian, F. Zhang, X. Xie, and G. Sun, "Restaurant survival analysis with heterogeneous information," in *Proceedings of the 26th International Conference on World Wide Web Companion (WWW Companion)*, 2017.
- [35] W.-T. Chu and Y.-L. Tsai, "A hybrid recommendation system considering visual information for predicting favorite restaurants," *World Wide Web Journal (WWWJ)*, 2017.
- [36] F. Zhang, N. J. Yuan, K. Zheng, D. Lian, X. Xie, and Y. Rui, "Exploiting dining preference for restaurant recommendation," in *Proceedings of the 25th International Conference on World Wide Web (WWW)*, 2016.
- [37] S. Bakhshi, P. Kanuparth, and E. Gilbert, "Demographics, weather and online reviews: A study of restaurant recommendations," in *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, 2014.
- [38] Y. Fu, B. Liu, Y. Ge, Z. Yao, and H. Xiong, "User preference learning with multiple information fusion for restaurant recommendation," in *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, 2014.
- [39] S. Auty, "Consumer choice and segmentation in the restaurant industry," *The Service Industries Journal*, 1992.
- [40] M. A. Clark and R. C. Wood, "Consumer loyalty in the restaurant industry: a preliminary exploration of the issues," *International Journal of Contemporary Hospitality Management*, 1998.