

Optimizing Email Marketing Campaigns in the Airline Industry using Knowledge Graph Embeddings

Amine Dadoun
EURECOM, Sophia Antipolis, France
Amadeus SAS, Biot, France
amine.dadoun@eurecom.fr

Raphaël Troncy
EURECOM, Sophia Antipolis, France
raphael.troncy@eurecom.fr

Michael Defoin Platel
Amadeus SAS, Biot, France
michael.defoinplatel@amadeus.com

Riccardo Petitti
Amadeus SAS, Biot, France
riccardo.petitti@amadeus.com

Gerardo Ayala Solano
Amadeus SAS, Biot, France
gerardo.ayalasolano@amadeus.com

Abstract

The use of email marketing campaigns in e-commerce is an essential aspect for driving sales and establishing or maintaining good customer service. However, ensuring their success in the airline industry is challenging, and the risk is high that the marketing campaign content is not suited to the needs of specific travelers. The traveler may rapidly feel spammed and the offer content promoted via emails may even be no longer accessed. Consequently, the use of personalised recommender systems for email marketing campaigns becomes essential to build user loyalty and to increase offer conversion. Recently, the use of knowledge graphs has proven to be an effective way to recommend products. In this work, we propose an approach that leverages knowledge graph embeddings to better target the right audience in email marketing campaigns for airline products recommendation. We conduct extensive experiments to compare our approach with the currently in-production rule-based system used by airline marketers and a supervised machine learning model based on handcrafted features as another baseline. The results suggest that the use of knowledge graph embeddings is the most effective approach.

A Introduction

With the digital transformation of retail commerce from physical sale points to virtual stores, recommender systems have shown their value in easing the search and purchase decision process of customers who are facing an ever increasing amount of products [1]. In order to embrace e-commerce techniques and boost their revenues, some airlines are using the so-called Amadeus Anytime Merchandizing (AAM) Notification System¹ which is an IT solution that allows airline marketers to effectively define, deploy, monitor and adjust airline email marketing

Copyright 2021 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

¹<https://amadeus.com/en/portfolio/airlines/anytime-merchandising>

campaigns sent to the travelers in real-time. Customized notifications can be defined and sent to travelers, after booking a flight, to suggest them additional services to buy (e.g. extra baggage, specific meal, preferred seat) called ancillaries. The solution acts as a bridge between the travelers retailing touchpoints and the airline’s service and delivery system. As shown on the left side of Figure 1, when using this solution, the airline marketer can choose the appropriate time **when** to send the notification (e.g. 5 days before departure), **what** product to recommend (e.g. Leg Space Seat), **how** to send the offer (e.g. via an Email) and to **whom** this offer should be sent (matching targeting criteria).

Despite all the functionalities included in the AAM Notification System, it is difficult for an airline to find the optimal audience to target for a given offer. We conducted an analysis of historical sales triggered by some notification campaigns during the period 14 May 2019 - 17 Dec 2019 ran by one of our partner airlines and we observed a poor conversion of the notification offers (Section C.3). This is partly due to the challenging decision-making process that an airline marketer faces when it comes to deciding which values (belonging to broad ranges) are appropriate for the criteria to be used (e.g. sending time, flight itineraries, flight departure point, etc.). Targeting customers with unsolicited notifications can be counter-productive and lead to adversarial effects on customer loyalty if done incorrectly. It is therefore critical to identify the customers that we expect to react positively to an advertised service in order to avoid spamming them with non-personalized emails. This problem can be seen as an inverse recommendation scenario, i.e. recommending a user to an item.

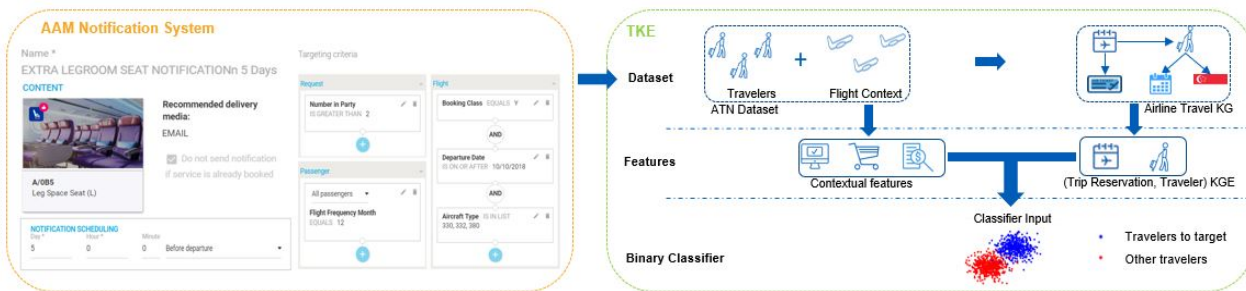


Figure 1: On the left side: AAM Notification System. On the right side: Flowchart of our proposed TKE framework. Notification dataset used in this study is generated from the AAM Notification system. Contextual features include booking context (e.g. number of passengers, date of departure, etc.), notification information (e.g. media used to send the notification, time of notification, etc.).

Inspired by recent works that have illustrated the effectiveness of using knowledge graph embeddings [2–4] and gradient boosting algorithms [5, 6] for item recommendation, we propose the **Travel Knowledge Graph Embeddings for email marketing campaigns (TKE)** framework to better target the audience for a service the airline wishes to recommend through email marketing campaigns (Figure 1). In this paper, we made the following contributions:

1. We design and develop a Knowledge Graph using Semantic Web technologies to represent the travelers past trips as well as to semantically enrich airline products.
2. We learn vector representations of travel entities via knowledge graph embedding algorithms and we leverage gradient boosting algorithms to compute prediction scores to better target the audience in email marketing campaigns.
3. We perform an empirical comparison of our approach with the current in-production rule-based system as well as with an hypothetical classical machine learning approach using handcrafted features on a real-world production dataset.

The paper is organized as follows. Section B provides a literature review of the related work. Section C introduces some preliminary concepts, the recommendation problem, the dataset and the knowledge graph used to conduct this work. In Section D, we describe both our TKE approach as well as a machine learning model using handcrafted features as another baseline. In Section E, we present the experiments we carried out and we demonstrate the effectiveness of our approach. Finally, in Section F, we give some conclusions and we outline some future works.

B Related Work

This section provides a literature review of the research conducted on email marketing campaigns in the e-commerce and tourism domains. We also present state-of-the-art methods of knowledge graph embedding algorithms and the use of gradient boosting algorithms for recommender systems.

B.1 Email Marketing Campaigns

Emails allow marketers to send messages to their customers at very low cost. They generally generate faster responses and create an opportunity for interactive communication with customers [7]. In [8], the authors analyze 70 randomized field experiments and find that email promotions not only increase customers' average purchase spending during the promotion window but also carry over to the week after the promotion expires. In our study, we will focus on personalized email marketing campaigns. In [9], the authors performed an empirical comparison of supervised machine learning models based on decision tree and logistic regression algorithms in order to improve the open rate and conversion rate of email marketing campaigns. In [10], the authors propose to use transactional features and instant messaging metadata to train a boosting tree regression algorithm to timely anticipate the needs of consumers in order to increase their level of engagement as well as the rate at which they repurchase products.

In the tourism domain, even if widely used, limited research is conducted on email marketing campaigns. In [11], the authors found that customers' favourite emails contain special offers, discounts and coupons as well as real-time communication tools. When customers perceived these emails as meeting their personal preferences, they developed a strong relationship with the sender. In [12], the authors demonstrated that the personalization, interactivity and price were important predictors of the possibility of revisiting the same accommodation. To the best of our knowledge, we are the first to propose a supervised machine learning approach to enable personalized email marketing in the airline travel industry.

B.2 Knowledge Graph Embeddings

A knowledge graph embedding (KG embedding) is a representation of a knowledge graph (KG) component into a continuous vector space. The objective of learning those embeddings is to ease the manipulation of graph components (entities, relations) for prediction tasks such as entity classification, link prediction or recommender systems.

In [13], the authors classified the learning algorithms into two main categories namely translational distance models and semantic matching models.

For the first category, TransE [14] is often mentioned as the most used translational distance model. TransE represents both entities and relations vectors in the same space R^d . Given a fact (h, r, t) , the relation is interpreted as a translation vector r so that the embedded entities h (head) and t (tail) can be connected by r with low error, i.e., $h + r \approx t$ when (h, r, t) holds. TransH [15] introduces relation-specific hyperplanes, each relation r being represented on a hyperplane as w_r its normal vector. TransR [16] follows the same idea as TransH, but instead of projecting the relations into a hyperplane, it proposes to create a specific space per relation.

In the second category, semantic matching models exploit similarity-based scoring functions. In [17], the authors associate each entity with a vector to capture its latent semantics. Each relation is represented as a matrix that models pairwise interactions between latent factors. The score of a fact (h, r, t) is defined by a bi-linear scoring function minimized through tensor factorization based on ALS optimization technique.

Palumbo et al. [18] propose to use property-specific KG embeddings generated from node2vec algorithm [19] in order to compute relatedness scores between items and users. Similarly, we propose to use translational distance and semantic matching models to generate embeddings and to use them as latent features of a gradient boosting algorithm.

B.3 Gradient Boosting Algorithms

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weaker prediction models.

In [5], the authors used multiple additive Regression Trees (Dart) which is an ensemble model that uses boosted regression trees and handles the overspecialization. Their approach for online accommodation recommendation ranked 1st in the famous RecSys 2019 Challenge². In [6], the authors used XGBoost [20] which is an implementation of gradient boosting decision trees to predict tweet engagement, they have intensively used exploratory data analysis to extract and compute relevant features to feed XGBoost algorithm. Their solution ranked 1st in the RecSys 2020 Challenge³. In our work, we use XGBoost algorithm as a supervised machine learning algorithm to better target the audience in email marketing campaigns based on the computed embeddings and contextual features (Figure 1).

C Data & Problem Formulation

In this section, we first provide definitions of some useful concepts. Then, we define the problem and research questions addressed in this work. Finally, we present the dataset and the design of the knowledge graph used for the experiments.

C.1 Preliminaries

We define a Knowledge Graph (KG) similarly to what is done in [21].

Definition 1: A knowledge graph is a set $K = (E, R, O)$, where E is the set of entities, $R \subset E \times \Gamma \times E$ is a set of typed relations between entities and O is an ontology. The ontology O defines the set of relation types ('properties') Γ .

Definition 2: Travelers are a subset of the entities of the KG, $t \in T \subset E$.

Definition 3: Ancillaries are all products offered by the airline beyond air tickets. They can be flight-related (e.g. extra baggage, preferred seat, etc.) or standalone (e.g. lounge access) services. Ancillaries are advertised in the email marketing campaigns. Ancillaries are a subset of the entities of the KG, $a \in A \subset E$.

Definition 4: A notification campaign is a set of notifications sent to an audience of travelers within a given period of time and under some criteria to recommend an ancillary product.

²<https://recsys.acm.org/recsys19/challenge/>

³<https://recsys.acm.org/recsys20/challenge/>

Definition 5: We define the conversion rate of a notification campaign as follows:

$$CR = \frac{1}{N_o} \sum_{i=1}^{N_o} hit(N_i) \quad (16)$$

where N_o is the number of notifications sent through the notification campaign, and $hit(N_i)$ is equal to 1 if the notification N_i triggers a purchase. In our work, we focus on optimizing the conversion rate.

C.2 Problem Formulation

Given a notification campaign aimed at a large audience of travelers who have already booked a flight in a given context, we aim to target the relevant travelers among all the travelers that the notifications will reach. More specifically, we address the following research questions:

RQ 1: How to extract the relevant sample of travelers to target for a given notification campaign?

RQ 2: How does a supervised machine learning approach perform compared to a rule-based approach to target the relevant audience for a notification campaign?

RQ 3: How does the use of KG embeddings compare to the use of handcrafted features as input of a supervised machine learning model trained to target the relevant audience for a notification campaign?

C.3 Notification Campaign Analysis

We analyzed three notification campaigns summing up to approximately 8.2 million notifications sent by one of our partner airlines to its travelers between 14 May 2019 and 17 December 2019 in order to understand the behavior of travelers in response to the notification campaigns, and to compute the conversion rates of these notification campaigns.

As shown in Table 16, there are three different types of ancillaries that are advertised in three notification campaigns. By analyzing airline sales data over the same period, we can see that only 3 out of 34 different types of purchased ancillaries were offered in the notification campaigns. This shows an untapped sales potential. Moreover, we observed that 50% of sales triggered by a notification happens on the same day (< 24 hours) after receiving the notification. This demonstrates the effect of a notification on the purchases.

Table 16: Conversion rates of notification campaigns: rule-based approach

Notification Campaign	Notification time	Date Range	Number of Notifications	Sales	CR
Extra leg room seat	5 days before Departure	19 May - 23 December 2019	~355 K	~2.8K	0.8%
Prepaid baggage	2 days before Departure	14 May - 17 December 2019	~7.5 M	~11K	0.15%
Lounge	Right after air ticket purchase	16 October - 17 December 2019	~338 K	104	0.03%
All Notifications	-	-	-	~13.8K	0.18%

While the prepaid baggage notification campaign was aimed at all travelers who booked a flight during the period indicated in Table 16, the notification campaigns for Lounge access and Extra leg room seat contain a number of filtering criteria that explain the large discrepancy in terms of the number of notifications sent out. Indeed, for these two notification campaigns, the airline marketer chose to send the notification to a quite restrictive audience by combining a number of criteria (fare family, aircraft type, no chargeable seat in their booking, etc.)

C.4 Airline Travel Notification Dataset

We conducted experiments on a real-world production dataset of bookings from the T-DNA database⁴. Each booking contains one or several air ticket purchases, and is stored using Passenger Name Record (PNR) information. The PNR is created at reservation time by airline reservation system and contains information about the purchased air ticket (e.g. travel itinerary, payment information), traveler demographics and ancillaries information if purchased (comprised in the EMD ticket). The considered dataset contains approximately 2.33 million bookings for approximately 2.85 million unique travelers.

The Airline Travel Notification (ATN) dataset is produced by joining the notification dataset and the historical bookings dataset from T-DNA. This dataset contains information about the shopping and booking context (e.g. search date, number of passenger, departure date, etc.) and information about travelers (e.g. demographics and loyalty membership information). In total, the dataset contains 42 columns and ~ 8.2 million rows. For our experiments, the dataset was broken down into three different sub-datasets that correspond to the three different notification campaigns (Table 16).

C.5 Airline Travel Knowledge Graph

The Airline Travel KG is constructed from the T-DNA database. We develop an ontology which is available in the Turtle format at https://gitlab.eurecom.fr/amadeus/tke4rec/-/blob/master/ontology/ams_ontology.ttl. To design the KG, we have defined 7 classes corresponding to top level entities and based on the various tables available in the T-DNA database:

- **Traveler:** A traveler is identified by a T-DNA id. A traveler has a booking history (PNRs) that contain purchase history (air tickets, EMD tickets). An instance of traveler is a `schema:Person`⁵.
- **Trip Reservation:** A trip reservation (PNR) represents the booking of all travelers contained in the PNR. It contains information such as the number of passengers, the destination, etc.
- **Journey:** A journey is linked to a trip reservation. Each journey has a stay duration, departure and arrival airports.
- **Air Ticket:** An air ticket is contained in a PNR and contains flight and transactional information. A PNR can have multiple air tickets because of the different flight legs (e.g. Nice-Paris, Paris-New York) or/and the number of passengers.
- **EMD Ticket:** An Electronic Miscellaneous Document (EMD) ticket is linked to an air ticket. It contains information on the ancillary purchased by the traveler (e.g. ancillary type, ancillary price, etc.).
- **Ancillary:** An ancillary is a service purchased by a traveler (associated to a flight) in addition to the air ticket. It is identified by a sub-code (RFISC), labelled by a commercial name, defined by ATPCO⁶. It belongs to a group of ancillaries (Group, RFIC). We propose to model the different ancillaries as SKOS concepts and we create an ancillary thesaurus represented as a concept scheme.
- **Airport:** It represents the airport where the traveler travels to. An airport serves one or several cities.

The KG used to tackle our use case contains 41 different properties (Figure 2), ~ 80 million edges and ~ 9 million nodes.

We present in Table 17 some statistics about the Airline Travel KG. In Figure 3, an excerpt of the KG is

⁴T-DNA: Traveler DNA is a database that contains bookings of travelers over a dozen of airlines

⁵The prefix `schema` is used for concepts defined by <https://schema.org>

⁶ATPCO Ancillary description: <https://www.atpco.net/resource/optional-services-industry-sub-codes>

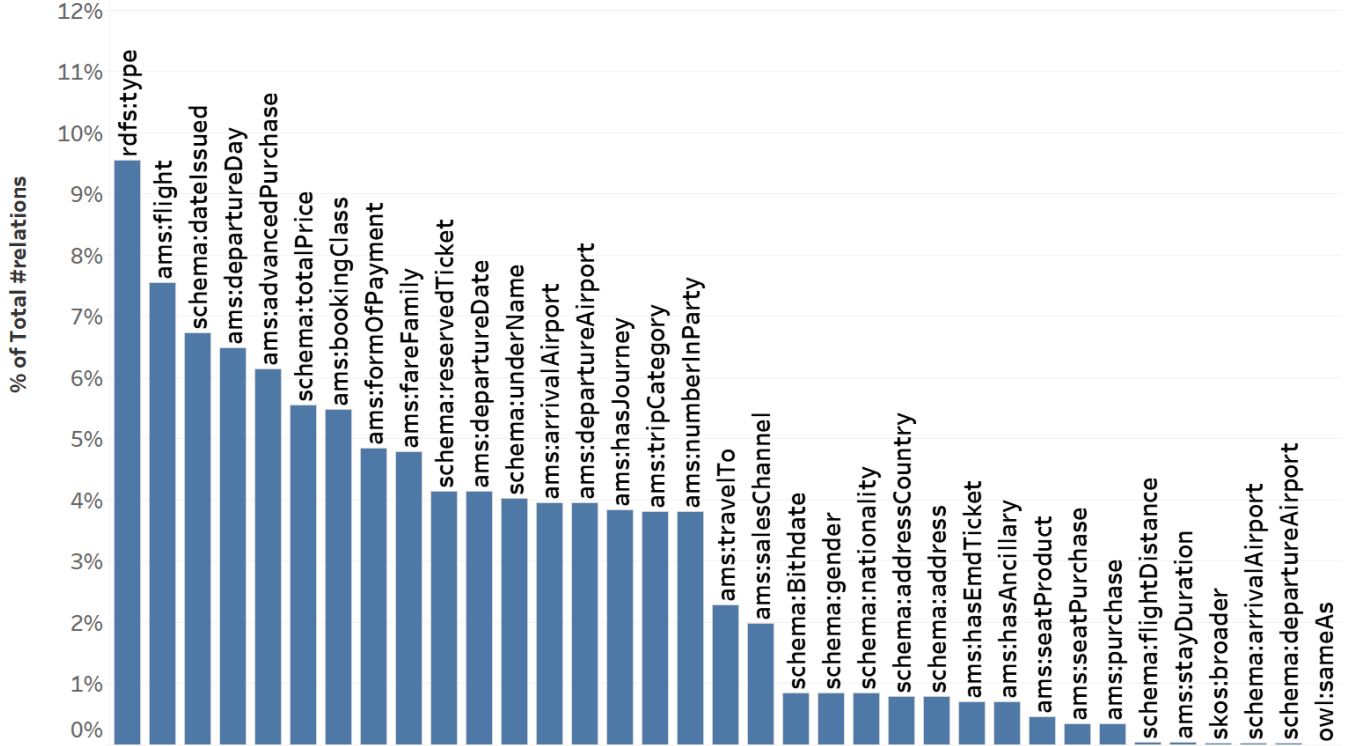


Figure 2: Distribution of #relations of properties in the Airline Travel KG. All prefixes can be found in the ontology definition.

Table 17: Statistics of subgraphs

Subgraph	#Edges	#Nodes	#travelers	#PNRs
Extra leg room seat	7M	800K	67K	205K
Prepaid baggage	64M	7.6M	572K	2.2M
Lounge	6.7M	789K	42K	203K

depicted, where a Malaysian traveler identified by T21354, born on "1988-05-05" has booked a one way flight for two people from Kuala Lumpur to Melbourne. The EMD ticket identified by 23143 and linked to the air ticket 21563 represents the purchase of an ancillary (a seat).

D Approach

Our proposed framework TKE can be seen as a two-stage approach as presented in Figure 1. In the first stage, we extract contextual features from the ATN dataset and compute KG embeddings of travelers and trip reservations from the Airline Travel KG. In the second stage, contextual features and KG embeddings are used as input of an XGBoost classifier in order to predict, for a given user, whether the notification should be sent or not.

We also propose a supervised machine learning model that includes as input contextual features and additional handcrafted travelers' features that capture travelers' preferences which we think could be particularly significant for model accuracy (hypothesis proven in the ablation study) as another baseline to compare with. We describe below the handcrafted travelers' features as well as the KG embeddings used in TKE:

- **Handcrafted travelers' features:** Features designed to capture travelers' preferences for ancillaries, destinations, points of sale, etc. and also customer lifetime value (Section D.1).

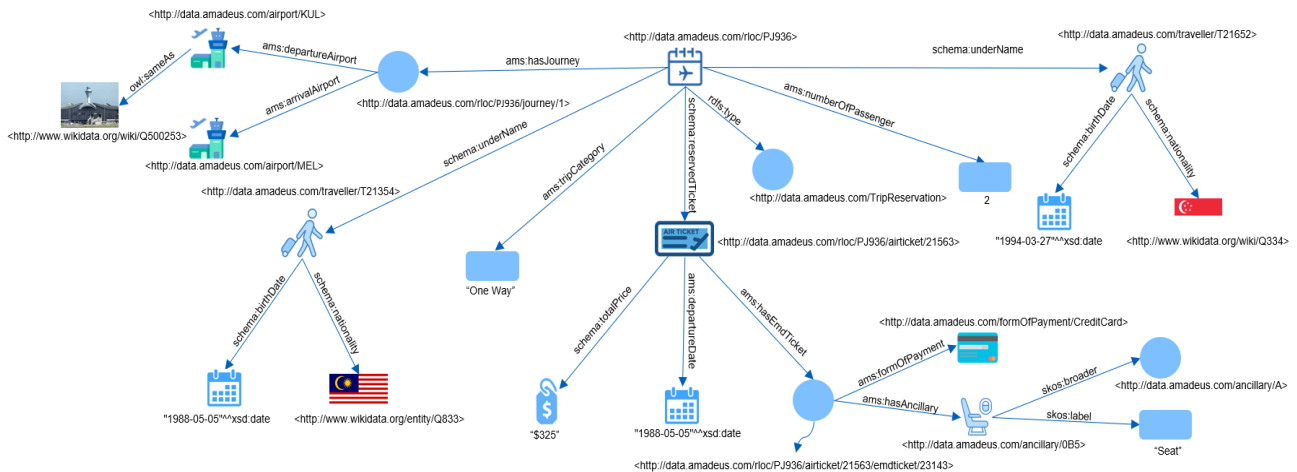


Figure 3: Excerpt of the knowledge graph representing the travelers included in a Trip reservation through the property `schema:underName`, as well as other properties and relations to other entities. Literals are represented in blue rectangle, whereas other entities are represented in blue circle. In this depiction, some properties that links travelers, trip reservations, air tickets and emd tickets are represented as an example, but more properties are included in the graph.

- **Knowledge graph embeddings:** Latent features representation of travelers and trip reservations computed based on KG embedding algorithms such as TransE [14] (Section D.2).

D.1 Handcrafted Travelers Features

In this section, we present the handcrafted travelers features designed to compare their use with that of KG embeddings. We compute several features based on travelers purchase history, such as preferred ancillary, preferred destination, etc. We list below the features computed that leverage travelers' history:

- *Bookings count:* Number of bookings already purchased by the traveler with the airline.
- *Average flight revenue:* The average booking price tickets for all historical bookings of the traveler.
- *Preferred ancillary:* This feature corresponds to the most sold ancillary to the traveler.
- *Preferred destination:* This feature corresponds to the most visited destination (airport) by the traveler.
- *Preferred seat characteristic:* This feature represents the seat characteristic that is the most purchased by the traveler. There are three types of seat characteristic namely Upper deck, Exit Row, Leg Space.
- *Extra leg room seat already purchase:* For each seat characteristic, we create a binary feature that represents if a traveler has already purchased an Extra leg room seat or not.
- *Seat sales count:* This feature represents the number of times a seat has been purchased by the traveler.
- *Prepaid baggage sales count:* This feature represents the number of times a prepaid baggage has been purchased by the traveler.
- *Lounge sales count:* This feature represents the number of times a lounge access has been purchased by the traveler.

- *Notification response rate*: This feature is equal to the number of sales divided by the number of notifications sent to the traveler (regardless of the recommended service).

The handcrafted features and the features available in the ATN dataset are used as input of a gradient boosting decision tree classifier. We use the official implementation⁷ of XGBoost [20] to train a binary classifier to predict if the notified travelers will convert or not. In Section E, we give more details about the hyper-parameters used in XGBoost.

D.2 TKE Features

In this section, we explain how the KG embeddings used in TKE are computed. We use translational distance models to compute travelers and trip reservations embeddings as shown in Figure 1. More formally, we learn the KG embeddings based on a link prediction task, where some links of ancillary purchases and seat products are hidden in the training set, and put in the test set. Translational distance models are trained under the closed world assumption [13] using a pairwise loss that penalizes negative instances. More concretely, ancillaries that were not purchased by a traveler are considered as negative instances under the closed world assumption. Translational distance models are evaluated using ranking metrics such as hit rate or mean reciprocal rank. Hence, these models will return a high similarity score (low euclidean distance) for the ancillaries that are close in the graph embedding space to the embeddings of the ancillaries historically purchased by the travelers. As an example, we obtain a hit rate of ~ 0.42 with the TransE algorithm on the Airline Travel KG. In addition to translational distance models, we implemented a single-hidden multi-layer perceptron (MLP) as proposed in [22], where each relation (as well as entity) is associated with a single vector. More specifically, given a fact (h, r, t), the vector embeddings of h, r, and t are concatenated in the input layer, and mapped to a non-linear hidden layer. The score is then generated by a linear output layer. The generated embeddings are used as input of XGBoost classifier in addition to the contextual features as shown in Figure 1. We carry out a thorough empirical comparison of the aforementioned KG embedding algorithms and select the KG embeddings that allow the classifier to predict with the highest accuracy.

E Experiments

The objective of the experiments is to compare the use of handcrafted features (a) with the use of KG embeddings (b). (a) helps in interpreting the results and predictions obtained by the algorithm, while (b) lacks interpretation (latent features), but is easier to compute and maintain. We publish our code as open source in order to ease reproducibility⁸.

E.1 Experimental Setup

In this section, we present the different settings and the evaluation protocol (evaluation metrics and split of the dataset) used to conduct the experiments.

Dataset: We experiment both approaches (a) and (b) with the three datasets presented in Table 16. We use the Airline Travel KG presented in Section C.5 to generate the KG embeddings useful for our main approach TKE.

Training & Test Sets: The three datasets corresponding to the three notification campaigns are split using the same strategy. Each dataset is sorted temporally, and 80% of the first rows of each dataset are used as training/validation sets. We use a cross-fold validation to train and validate all models ($k=5$, a split of 80% for training and 20% for validation). The remaining 20% are used as test set to evaluate the model. The split between training and validation set is performed randomly in order to avoid a seasonality effect that is usually occurring in

⁷XGBoost:<https://XGBoost.readthedocs.io/>

⁸<https://gitlab.eurecom.fr/amadeus/tke4rec>

Table 18: Evaluation results of the different approaches. (a) represents the results of XGBoost for different inputs; (b) represents the results of the TKE approach for different KG embedding algorithms. The average standard deviation (by varying the seed when splitting the dataset) of each metric is as follows: $AUC - ROC : \pm 0.02$, $TPR : \pm 3\%$, $TNR : \pm 2\%$, $CR : \pm 0.1\%$

Model	Extra leg room seat				Prepaid baggage				Lounge			
	AUC-ROC	TPR	TNR	CR	AUC-ROC	TPR	TNR	CR	AUC-ROC	TPR	TNR	CR
Rule-based	-	-	-	0.8%	-	-	-	0.15%	-	-	-	0.03%
(a) ADS	0.75	78%	58%	2.2%	0.83	80%	71%	0.38%	0.76	80%	62%	0.18%
(a) HDS	0.79	81%	60%	2.37%	0.85	82%	74%	0.4%	0.84	86%	67%	0.22%
(a) AHDS	0.83	85%	65%	2.8%	0.88	86%	74%	0.56%	0.89	88%	65%	0.36%
(b) TransE	0.85	86%	69%	3.1%	0.91	92%	65%	0.6%	0.90	89%	78%	0.35%
(b) TransH	0.84	85%	67%	3%	0.90	91%	65%	0.59%	0.95	96%	85%	0.59%
(b) TransR	0.84	85%	67%	2.9%	0.90	91%	65%	0.6%	0.92	92%	80%	0.52%
(b) MLP	0.87	88%	69%	3.2%	0.92	94%	65%	0.62%	0.91	90%	81%	0.56%

the travel industry. KG embedding algorithms are often designed to solve a link prediction task. We consider it is appropriate to split the KG by removing some edges that are included in the set of properties that link travelers with ancillaries and consider them as test sets, in order to evaluate the quality of the computed embeddings.

Evaluation metrics: The output of our approach is the probability of purchasing the recommended ancillary a included in the notification N :

$$P(\text{purchase} = a|N) = P(\text{purchase}|Context, TE, RE) \quad (17)$$

where, TE and RE are the Traveler and Trip reservation embeddings.

To evaluate and compare, the different approaches implemented, we used the conversion rate defined at definition 5 and the three metrics defined as follows:

- **TPR:** The true positive rate is the percentage of correct positive predictions. It represents the ratio of travelers that the algorithm suggests to send the notification and effectively purchase the ancillary. TPR is defined as follows:

$$TPR = \frac{TP}{(TP + FN)} \quad (18)$$

- **TNR:** The true negative rate is the percentage of correct negative predictions. It represents the ratio of travelers that the algorithm suggest to not send the notification and effectively do not purchase the ancillary. TNR is defined as follows:

$$TNR = \frac{TN}{(TN + FP)} \quad (19)$$

- **ROC-AUC:** The area under ROC curve (FPR, TPR) helps to choose what is the optimal probability threshold that maximizes the CR and TPR and is defined as follows:

$$ROC-AUC = \int_0^1 TPR d(FPR) \quad (20)$$

where, $FPR = 1 - TNR$ is the false positive rate

It is noteworthy that the conversion rate was measured offline as well as all the metrics based on the test set. According to equation 16, N_o represents the number of predicted positives and each hit hit_i corresponds to a true positive prediction.

Implementation Framework & Parameter Settings: For KG embedding algorithms, we use the deep learning framework pytorch⁹ to implement MLP [22] and the library pykg2vec [23] for all the other KG embedding algorithms. The hyper-parameters of all the models were tuned using a combination of random-search and grid-search algorithms. We apply grid-search algorithm on the implemented algorithms using the following values: the embedding size $k \in \{32, 64, 96, 128, 256\}$, the batch size $\in \{128, 256, 512, 1024\}$, the number of epochs $\in \{50, 100, 200\}$, the learning rate $lr \in \{0.001, 0.003, 0.01, 0.03, 0.1, 0.3\}$ and negative samples $N_s \in [2, 10]$ for MLP algorithm. We also optimize the following hyper-parameters of XGBoost classifier: the max depth of a tree $\in [5, 50]$, the number of trees $\in [10, 100]$, the sub-sample of each tree $\in [0.65, 0.85]$ and the col-sample of each tree $\in [0.65, 0.85]$. In addition to these hyper-parameters, we compute a weighted score (ratio of number of negative class to the positive class) that we use in XGBoost to approach the problem as a cost-sensitive learning problem due to the high class imbalance between positive (purchase) and negative (no purchase) classes (Table 16).

E.2 Results and Discussion

In this section, we discuss the results obtained from the experiments. Results of the experiments conducted are presented in table 18. TPR, TNR and ROC-AUC metrics are not provided for the rule-based approach implemented in AAM Notification System. The reason behind this is that the dataset used in the experiments is generated by the AAM notification system, which is different from the original dataset that contains all travelers used for the rule-based approach to identify the travelers matching the targeting criteria.

Ablation Study: Table 18 shows that using the features from the ATN dataset in addition to the travelers handcrafted features (AHDS: ATN + Handcrafted features) as input of XGBoost performs better than using only one of them (ADS: ATN features or HDS: Handcrafted features) as input for all notification campaigns. We also observe that using only travelers handcrafted features as input information of XGBoost gives better results than using the entire ATN dataset for all notification campaigns. We compute the most important features of the model (a) AHDS for each notification campaign and we report below the three most important ones with their respective information gain:

- Extra Leg Room Seat: *{Preferred Seat Characteristic: 0.31, Preferred ancillary: 0.12, Ticket amount: 0.08}*.
- Prepaid Baggage: *{Preferred destination: 0.21, Destination: 0.12, Prepaid Baggage sales Frequency: 0.10}*.
- Lounge: *{Average Flight Revenue: 0.22, Destination: 0.20, Age: 0.15}*.

Knowledge graph embeddings: We observe in Table 18 that using KG embeddings (concatenation of traveler and reservation KG embeddings) with contextual features as input of XGBoost performs better than using travelers handcrafted features regardless of the algorithm used to compute the embeddings or the notification campaign. Moreover, KG embeddings computed from **MLP** shows to perform better than KG embeddings computed from translational distance models except for the lounge notification campaign, where the use of KG embeddings computed from **TransH** model gives the best results.

F Conclusions and Future Work

In this work, we have presented a two stage approach to address the problem of audience targeting for email marketing campaigns: first, we compute KG embeddings of travelers and reservations; second, we use these

⁹Pytorch:<https://pytorch.org/>

embeddings in addition to contextual features as input of a XGBoost classifier to learn what is the relevant audience to target for a given notification campaign. We conducted several experiments to address our research questions:

RQ 1: The results of the experiments presented in Table 18 show that extracting the relevant audience for a given notification campaign is not an easy task. Indeed, despite the fact that the conversion rate increases significantly with our approach, it remains relatively small. However, thanks to our approach, notification campaigns are better targeted and we manage to avoid recommending an unsuitable service to at least 65% of passengers.

RQ 2: Experiments have shown that the handcrafted features based supervised machine learning approach (a) gives better results than the rule-based one. Indeed, in Table 18, we can observe that the conversion rate is multiplied by more than 3 for Extra Leg Room Seat, almost 4 for Prepaid Baggage, and 12 for Lounge. Hence, we prove the benefit of using supervised machine learning over a simpler rule-based approach while it is the currently adopted mechanism used by airline marketers. It should be noted that the list of possible criteria available in AAM Notification System (Figure 1) is the same as the list of features used in the supervised machine learning approach.

RQ 3: Experiments show that regardless of the KG embedding algorithm tested, the KG embedding approach is better than the handcrafted features approach. This is very interesting from a scientific point of view, as it shows the added value of having a KG in the travel domain that could be used not only for ancillary recommendation task, but also other recommendation tasks (e.g. Trip recommendation) as the same KG embeddings could be used. It is worth noticing that when dealing with a cold-start problem (new user or item) for on-line usability, a rule-based approach is more appropriate.

Finally, as future work, we expect to tackle the task of personalized ancillary ranking in email marketing campaigns. More specifically, the goal of our future work will be to answer what is the most appropriate service to recommend in a notification campaign. In addition to addressing and optimizing what to recommend to a traveler, it would be interesting to optimize the time when to send the notification as it is an important decision making factor, especially in the airline travel industry [1].

References

- [1] A. Dadoun, T. Fiig, M. Defoin-Platel, C. Landra, and R. Troncy, “How recommender systems can transform airline offer construction and retailing,” *Journal of Revenue and Pricing Management, Special issue on AI*, 2021.
- [2] E. Palumbo, G. Rizzo, R. Troncy, E. Baralis, M. Osella, and E. Ferro, “Translational models for item recommendation,” in *The Semantic Web: ESWC 2018 Satellite Events*, A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, and M. Alam, Eds. Cham: Springer International Publishing, 2018, pp. 478–490.
- [3] Z. Sun, J. Yang, J. Zhang, A. Bozzon, L.-K. Huang, and C. Xu, “Recurrent Knowledge Graph Embedding for Effective Recommendation,” in *12th ACM Conference on Recommender Systems*. New York, NY, USA: ACM, 2018, pp. 297–305.
- [4] E. Palumbo, D. Monti, G. Rizzo, R. Troncy, and E. Baralis, “entity2rec: Property-specific knowledge graph embeddings for item recommendation,” *Expert Syst. Appl.*, vol. 151, p. 113235, 2020.
- [5] P. Jankiewicz, L. Kyrashchuk, P. Sienkowski, and M. Wójcik, “Boosting algorithms for a session-based, context-aware recommender system in an online travel domain,” in *Proceedings of the Workshop on ACM Recommender Systems Challenge*, ser. RecSys Challenge ’19. New York, NY, USA: Association for Computing Machinery, 2019.

- [6] B. Schifferer, G. Titericz, C. Deotte, C. Henkel, K. Onodera, J. Liu, B. Tunguz, E. Oldridge, G. De Souza Pereira Moreira, and A. Erdem, “Gpu accelerated feature engineering and training for recommender systems,” in *Proceedings of the Recommender Systems Challenge 2020*, ser. RecSysChallenge ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 16–23. [Online]. Available: <https://doi.org/10.1145/3415959.3415996>
- [7] L. Chittenden and R. Rettie, “An evaluation of e-mail marketing and factors affecting response,” *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 11, pp. 203–217, 2003.
- [8] N. S. Sahni, D. Zou, and P. K. Chintagunta, “Do targeted discount offers serve as advertising? evidence from 70 field experiments,” *Manage. Sci.*, vol. 63, no. 8, pp. 2688–2705, Aug. 2017. [Online]. Available: <https://doi.org/10.1287/mnsc.2016.2450>
- [9] R. Abakouy, E. M. En-Naimi, and A. El Haddadi, “Classification and prediction based data mining algorithms to predict email marketing campaigns,” in *Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems*, ser. ICCWCS’17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: <https://doi.org/10.1145/3167486.3167520>
- [10] A. Deligiannis, C. Argyriou, and D. Kourtesis, “Predicting the optimal date and time to send personalized marketing messages to repeat buyers,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, 2020. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2020.0110413>
- [11] K. Y. Tam and S. Y. Ho, “Web personalization as a persuasion strategy: An elaboration likelihood model perspective,” *Info. Sys. Research*, vol. 16, no. 3, pp. 271–291, Sep. 2005. [Online]. Available: <https://doi.org/10.1287/isre.1050.0058>
- [12] K. Yang, J. H. Min, and K. Garza-Baker, “Post-stay email marketing implications for the hotel industry: Role of email features, attitude, revisit intention and leisure involvement level,” *Journal of Vacation Marketing*, vol. 25, no. 4, pp. 405–417, 2019.
- [13] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, Dec 2017.
- [14] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2013, pp. 2787–2795. [Online]. Available: <http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf>
- [15] Z. Wang, J. Zhang, J. Feng, and Z. Chen, “Knowledge graph embedding by translating on hyperplanes,” in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, ser. AAAI’14. Québec City, Québec, Canada: AAAI Press, 2014, pp. 1112–1119.
- [16] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, “Learning entity and relation embeddings for knowledge graph completion,” in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, ser. AAAI’15. Austin, Texas: AAAI Press, 2015, pp. 2181–2187.
- [17] M. Nickel, V. Tresp, and H.-P. Kriegel, “A three-way model for collective learning on multi-relational data,” in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ser. ICML’11. Madison, WI, USA: Omnipress, 2011, pp. 809–816.

- [18] E. Palumbo, G. Rizzo, and R. Troncy, “Entity2rec: Learning user-item relatedness from knowledge graphs for top-n item recommendation,” in *Eleventh ACM Conference on Recommender Systems*. New York, NY, USA: ACM, 2017, pp. 32–36.
- [19] A. Grover and J. Leskovec, “Node2vec: Scalable feature learning for networks,” in *22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016, pp. 855–864.
- [20] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794.
- [21] A. Dadoun, R. Troncy, O. Ratier, and R. Petitti, “Location Embeddings for Next Trip Recommendation,” in *9th International Workshop on Location on the Web (LocWeb)*. San Francisco, USA: Association for Computing Machinery, 2019, pp. 896–903.
- [22] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang, “Knowledge vault: A web-scale approach to probabilistic knowledge fusion,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 601–610. [Online]. Available: <https://doi.org/10.1145/2623330.2623623>
- [23] S.-Y. Yu, S. R. Chhetri, A. Canedo, P. Goyal, and M. A. A. Faruque, “Pykg2vec: A python library for knowledge graph embedding,” *Journal of Machine Learning Research*, vol. 22, no. 16, pp. 1–6, 2021. [Online]. Available: <http://jmlr.org/papers/v22/19-433.html>