



**HAL**  
open science

## Exploring eating behaviours modelling for user clustering

Sema Akkoyunlu, Cristina Manfredotti, Antoine Cornuéjols, Nicolas Darcel,  
Fabien Delaere

► **To cite this version:**

Sema Akkoyunlu, Cristina Manfredotti, Antoine Cornuéjols, Nicolas Darcel, Fabien Delaere. Exploring eating behaviours modelling for user clustering. HealthRecSys@RecSys 2018 colocated with ACM Recsys'18 (ACM Conference Series on Recommender Systems), Oct 2018, Vancouver, Canada. pp.46-51. hal-02480562

**HAL Id: hal-02480562**

**<https://hal.science/hal-02480562v1>**

Submitted on 16 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploring eating behaviours modelling for user clustering

Sema Akkoyunlu  
UMR MIA-Paris, AgroParisTech,  
INRA  
Université Paris-Saclay  
Paris, France  
sema.akkoyunlu@agroparistech.fr

Cristina Manfredotti  
UMR MIA-Paris, AgroParisTech,  
INRA  
Université Paris-Saclay  
Paris, France  
cristina.manfredotti@agroparistech.fr

Antoine Cornuéjols  
UMR MIA-Paris, AgroParisTech,  
INRA  
Université Paris-Saclay  
Paris, France  
antoine.cornuejols@agroparistech.fr

Nicolas Darcel  
UMR PNCA, AgroParisTech, INRA  
Université Paris-Saclay  
Paris, France  
nicolas.darcel@agroparistech.fr

Fabien Delaere  
Danone Nutricia Research  
Palaiseau, France  
fabien.delaere@danone.com

## ABSTRACT

Food based dietary guidelines are not fully adopted by consumers. One of the principal explanations for this failure is that they are too general and do not take into account eating habits. Experts in nutrition believe that providing personalized dietary recommendations via nutrition recommender system can help people improve their eating behaviours. Understanding eating habits is a keystone in order to build a context aware recommender system that delivers personalized dietary recommendations. As a step towards this goal, we propose a method for representing food consumptions based on Doc2Vec for discovering clusters of eating behaviours. We compare our method to the state of the art methods used in the nutrition community.

## CCS CONCEPTS

• **Information systems** → **Information extraction**; • **Human-centered computing** → **User models**;

## KEYWORDS

food recommender systems; user modelling; eating behaviours; Doc2Vec

### ACM Reference format:

Sema Akkoyunlu, Cristina Manfredotti, Antoine Cornuéjols, Nicolas Darcel, and Fabien Delaere. 2018. Exploring eating behaviours modelling for user clustering. In *Proceedings of Third International Workshop on Health Recommender Systems co-located with Twelfth ACM Conference on Recommender Systems, Vancouver, BC, Canada, October 6, 2018 (HealthRecSys'18)*, 6 pages.

## 1 INTRODUCTION

Most chronic diseases such as diabetes, obesity and cardiovascular diseases are correlated to unhealthy eating habits [25]. In order to help people to adopt healthier eating habits, public health agencies have created dietary guidelines targeted to the general population. These guidelines can be food based, for instance "eat at least 5 fruit or vegetable per day", "limit your consumption of salt"<sup>1</sup> or

nutrient based "XX gram of iron per day". However, the compliance to the guidelines are relatively low although the awareness about food based dietary guidelines is rather good [8]. Several causes contribute to this phenomenon: cultural and personal preferences, difficulty of implementing dietary changes, availability and price of food items [22]. One solution to this problem could be to provide a food recommender system able to take into account most of these causes. Early studies showed that web-based personalized interventions are more effective than standard public health advice for inducing compliance with healthy eating recommendations [6]. Moreover changing eating habits is challenging, thus food based recommendations should better be easy to follow [1]. But for recommendations to be practical, one should first understand consumers' eating behaviour.

In food related recommender systems, the recommended objects are recipes [4] [20], food items [5] or menus [3]. *Recipe recommendation systems* take advantage of users' past recipes ratings to propose recipes that they might like. *Menu based recommendation systems* combine meals that users showed preference for with nutritional constraints based on the nutritional requirements of users. *Food item based recommendation systems* are designed to learn the users' tastes for food items. Most of them use popular recommendation algorithms often based on matrix factorizations techniques which learn an embedding space for representing users and food items simultaneously. However, this representation does not take into account that food items are seldom consumed in isolation and that users' preferences for food items can change in response to the other food items consumed (i.e the dietary context) and to the context of consumption (e.g. eating croissant for breakfast is acceptable, but it is not for lunch). It seems necessary to take into account these aspects for increasing the efficacy of food item recommendation in real-life settings. Context-aware recommender systems seem therefore to be the appropriate approach. However, modelling the context is highly dependent on the domain at hand. It is thus necessary to first model eating behaviours and understand how the context impacts eating behaviours.

Several dietary assessment methods are available: the food frequency questionnaire (FFQ), 24-hour dietary recall (24HR) and food diaries. FFQ are easy to implement and cost-effective however, the

<sup>1</sup>HealthRecSys'18, October 6, 2018, Vancouver, BC, Canada. ©2018 Copyright for the individual papers remains with the authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

<sup>1</sup>[http://solidarites-sante.gouv.fr/IMG/pdf/PNNS\\_2011-2015.pdf](http://solidarites-sante.gouv.fr/IMG/pdf/PNNS_2011-2015.pdf)

User	ID_meal	Meals
Anna	$m_1$	coffee, cereals
	$m_2$	pasta, beef, fruits
	$m_3$	coffee
	$m_4$	rice, vegetable, fruits
Bob	$m_3$	coffee
	$m_5$	pizza, soda
	$m_3$	coffee
	$m_6$	pasta, soda
Christian	$m_7$	tea, cereals,
	$m_8$	pasta, vegetable
	$m_9$	tea, cereals, fruits
	$m_4$	rice, vegetable, fruits

**Table 1: Toy example of food consumption data (10 food items, 4 meals per user)**

questionnaire is tailored by research groups with a specific aim in mind. Besides, its accuracy is not enough for recommendation purposes. 24HR method is an interview that requires 30 minutes rather precise but one day of consumption per user is not sufficient in order to learn preferences. Food diaries are a prospective open-ended food consumption assessment method where consumers write down all the food items and beverages consumed over a specific time period [19]. Quite often, the time periods go from 3 to 7 consecutive days. The main advantages are that no interviewer is required, the whole process can be automatized adapted for recommendation purposes and provided several days of consumption, changes in diet can be captured. Throughout the paper, the toy food diaries dataset in Table 1 will be used to illustrate the user modelling methods.

Dietary behaviour is modelled using two main types of methods: theoretical ones and empirical ones [15]. Theoretical methods use dietary indexes developed by research groups or agencies in order to rank the healthiness of eating behaviours. Indexes are constructed based on the current knowledge in nutrition but can also include current dietary guidelines and recommendations which are usually generated from empirical research. However, Newby et al. [15] stress the fact that there can be conflicts when there is no scientific consensus about what a healthy behaviour is before analysis. It results in indexes that measure different definitions of a healthy behaviour. In empirical methods, there is no nutritional a priori about eating behaviours, i.e there is no definition about what a healthy behaviour is. Patterns are found with no nutritional a priori. We only focus on empirical ones as our goal is to learn dietary behaviours based on consumption data in an unsupervised way. In the literature, two methods stand out for discovering eating behaviours: clustering and factor analysis. Cluster analysis aims at discovering groups of behaviours, while factor analysis seeks the most relevant factors. Clustering may use factor analysis as a preprocessing step. Thus, the K-Means algorithm is often applied to the matrix of consumption of food items directly [17] or after dimension reduction using e.g. Principal Component Analysis (PCA) [21] or Non-Negative Matrix Factorisation [26].

To our knowledge, there is no comprehensive review about methods used for deriving empirically eating patterns [15]. Each study

works on its own dataset and, most of the time, only one method of dimension reduction is applied for deriving eating behaviours. There is no apparent gold standard method, but the existing literature seems to favour the use of PCA.

These methods are reductionist: they only consider food items alone. Nutrition experts argue that this reductionist perspective may not be efficient for recommendation purposes: deeper and more complex information are needed [23]. Opposed to the reductionist viewpoint, the holistic approach considers the diet as "a dynamic interaction of the parts of their synthesis" [7]. Food item interactions should accordingly be used for modelling eating behaviours.

One solution would be to consider dietary data in a meal-based form. Meal pattern analysis provides more details regarding the way people compose their meals [24] and could provide more insights for characterising eating behaviours. This approach takes into account the complexity of the diet and aims at overcoming the limitations of the study of foods in isolation [7]. A meal based approach for discovering eating behaviours was introduced by Woolhead et al. [24]. They used frequent itemsets to generate a generic meal classification. They derived 63 generic meals across all meal types and computed mean daily nutrient intakes associated to the generic meals. For each subject, mean daily intakes of energy percentage contribution of each generic meal type was computed. Then PCA was applied to discover eating behaviours. Authors themselves argue that this methodology induces a subjective classification. Besides, relying on frequent itemsets to code meals may overlook infrequent eating patterns at a population level but frequent at an individual level, discarding these patterns as noise. This shows the necessity of an adequate representation of meals.

Developing a food recommender system that takes into account the meals and their context, and not only food items, requires that two main challenges be met: (1) finding a proper meal description model in which distances between meals can be computed and (2) discovering an adequate way of aggregating several meals for computing distances between users in order to discover clusters of eating behaviours.

In this paper, our contribution is twofold: we propose a novel domain of application of word embedding to user profiling and we compare three approaches to describe eating behaviours. We propose a new approach to model meal representation by applying the Doc2Vec algorithm [10] in order to learn a meal embedding space. This allows, in turn, the use of a cosine similarity adapted to matrices to compute similarities between users and infers clusters of users. Moreover, in food based approaches, we compare the state in the art methods with Doc2Vec applied on users.

The rest of the paper is organised as follows. Section 2 describes methods for user modelling. Section 3 reports the results of our experiments on a real-world dataset. We discuss the results in Section 4 and we finally conclude in Section 5.

## 2 METHODS

### 2.1 Food-based methods

#### 2.1.1 State of the art methods.

In alimentation behaviour science, researchers work mostly on food items. They transform food consumption data into matrices where the columns correspond to the frequency or the quantity

of consumption of food items and the rows to users as shown in Figure 1. The next step consists in applying Principal Component

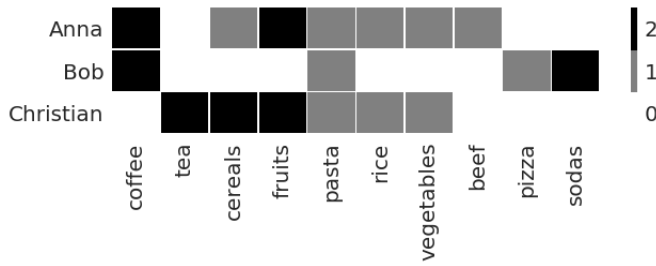


Figure 1: Matrix of consumption of the toy example

Analysis (PCA) or Non-Negative Matrix Factorization (NMF) [11]. PCA consists in finding a set of linearly independent variables, called principal components, that capture as much as possible the variance of the data points. NMF is similar to PCA but imposes a non-negativity constraint on the parameters of the model. This is found useful in many domains such as signal processing and recommender systems, because more amainable to interpretation by experts [12].

Clusters of eating behaviours are then discovered by applying K-Means algorithm on the result of PCA or NMF. In order to find the optimal number of clusters, a popular clustering evaluation metric is used, the silhouette coefficient [18].

2.1.2 Another food based method: applying Doc2Vec to users.

Word2Vec is a popular model for word embedding. Doc2Vec, proposed by [10] is an extension of Word2Vec: instead of learning word embeddings, the model learns distributed representations of arbitrarily large units of text such as sentences, paragraphs or documents. It was proposed in two flavours: DBOW (Distributed Bag Of Words) and DMPV (Distributed Memory version of Paragraph Vector). DBOW is simpler than DMPV as it does not take into account the order of the words when learning the embedding space. It is the version that is suited for our task as the order does not matter. Besides, empirical evaluations of Doc2Vec showed that DBOW performs better than DMPV [9].

The food based approach considers that a user is described by the frequency of consumption of single food items. Similarly, a user can be considered as a document where the food items eaten over a specific amount of time play the role of words.

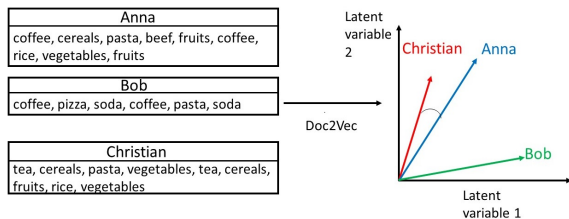


Figure 2: Application of Doc2Vec on user consumptions in the food based approach

Figure 2 is an illustration of what applying Doc2Vec algorithm on individual eating consumptions means. Individual documents of consumption are fed in the model. The result is an embedding space of users based on their eating consumptions which means that each user is described by a set of coordinates. Users are represented as vectors in this figure because similarity between users is computed with cosine similarity, a metric commonly used in document retrieval. It is basically the angle between two user vectors. At this stage of the method, we compute the similarity matrix of users. Our goal is then to cluster users according to their similarity.

Spectral clustering is a method that exploits similarity measures by considering data points as nodes of a weighted connected graph. Clusters are found by partitioning this graph based on the eigenvectors of the Laplacian matrix derived from the similarity matrix. Choosing the optimal number of clusters is often a problem for clustering algorithms. There are several heuristics adapted for spectral clustering. The heuristic advised by [13] is the eigengap heuristic. The optimal number of clusters  $k$  is the number such that the difference between the eigenvalues  $\lambda_{k+1} - \lambda_k$  is large. Justification for this procedure is provided in [13].

2.2 A novel meal based method using Doc2Vec

2.2.1 Learning an embedding space for meals.

Meals are defined as combinations of food items simultaneously consumed by one user at a single moment of consumption on one survey day. Meals are actually lists of food items. In the meal based approach, the objective is to be able to compute similarities between meals in order to compute similarities between users to derive clusters of users. However, it is not trivial to compute similarity between two meals, for example between  $\{pasta, beef, fruits\}$  and  $\{rice, vegetable, fruits\}$ .

A straightforward idea would be to define first a similarity between food items and then define a way to summarize those similarities to compute a similarity between meals. We observe two problems with this idea. First, there is no domain similarity measure between food items. One can use classification of food items as a proxy to a similarity measure. But there are lots of classification schemes in the literature. Second, this approach is against the philosophy of the holistic approach as it ignores interactions that may exist between food items.

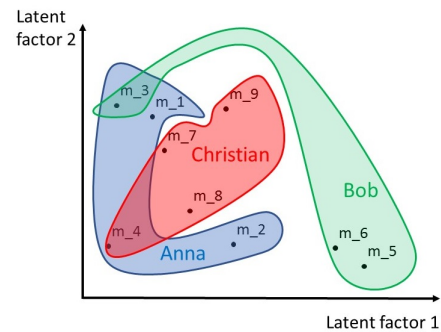


Figure 3: Application of Doc2Vec on meals in the meal based approach

An elegant way of learning such interactions is to learn an embedding space with Doc2Vec. Indeed, the embedding is learned in such way that similar meals are closer in the induced space as showed in Figure 3. Each user is now described by a matrix where the rows correspond to the meals and the columns to the coordinates of meals in the Doc2Vec induced space.

### 2.2.2 Computing distance between users.

Once the meal representation is learned, the challenge becomes one of computing a similarity between users. In our approach, this amounts to compute the similarity between two documents by taking into account the distances between sentences. Indeed, meals can be considered sentences of users who are documents. Mathematically speaking, this amounts to compute a similarity between matrices. Such a similarity was introduced in [14]. The authors of the paper proposed a cosine kernel in order to compute the similarity between the documents A and B in Equation 1:

$$\cos(A, B) = \frac{\langle A, B \rangle}{\|A\|_F \cdot \|B\|_F} \quad (1)$$

where  $\langle \cdot, \cdot \rangle$  is the Frobenius inner product and  $\|\cdot\|_F$  the Frobenius norm. Using the Frobenius inner product enables to compare the similarity of the sentences to determine the similarity of the documents. Let us denote  $s_A$  and  $s_B$  the number of sentences in document A and document B respectively. This formula implies that the cosinus similarity is computed between the first sentences of both documents then the second ones and so on until the  $\min(s_A, s_B)$ -th sentences. If one document is longer than the other one, the last sentences of the longer document are not taken into account for the similarity computation.

For eating behaviour modelling, this means that two consumers are similar if they eat similar meals at the same moment of the day on the same day. This is a rather strong assumption concerning eating behaviour modelling.

## 3 EXPERIMENTS

### 3.1 INCA2 dataset

The INCA2 dataset<sup>2</sup> consists of individual 7-day food records collected during 2006-2007 from 2,624 adult French consumers over several months in order to take into account seasonality. A close-ended list of 1,342 food items organized in 122 sub-groups and in 44 groups were used for coding the dietary records. Further detail about the survey methods can be found in [2]. We decide to work on sub-groups because the vocabulary is larger than when considering groups while having enough repetitions unlike when considering food items. We do not impose the number of clusters to be the same for all the methods as we want to see if the number of clusters that each method discovers is different, if the clusters are overlapping or not.

### 3.2 PCA and NMF on consumption data

The state of the art methods require the selection of two parameters: the number of components  $C$  of the reduction of dimensionality method and the number of clusters  $K$ . The number of clusters  $k$  is

<sup>2</sup><https://www.data.gouv.fr/fr/datasets/donnees-de-consommations-et-habitudes-alimentaires-de-letude-inca-2-3/>

determined by using an internal clustering evaluation score, the silhouette score. The optimal number of clusters is found when the silhouette score is maximised. For PCA and NMF, we vary the number of clusters between 2 and 30 and compute the silhouette score. The score is maximised for  $k = 9$ .

Loadings of factors of PCA and NMF can be give a hint about the new representation space of users. Figure 4 shows the loadings of factors of PCA according to food items. For ease of reading only food items whose absolute value of contribution to any factor is superior to 0.005 are displayed. NMF factors are shown in Figure 5. The food items are displayed if their contribution to any factor is superior to 0.3.

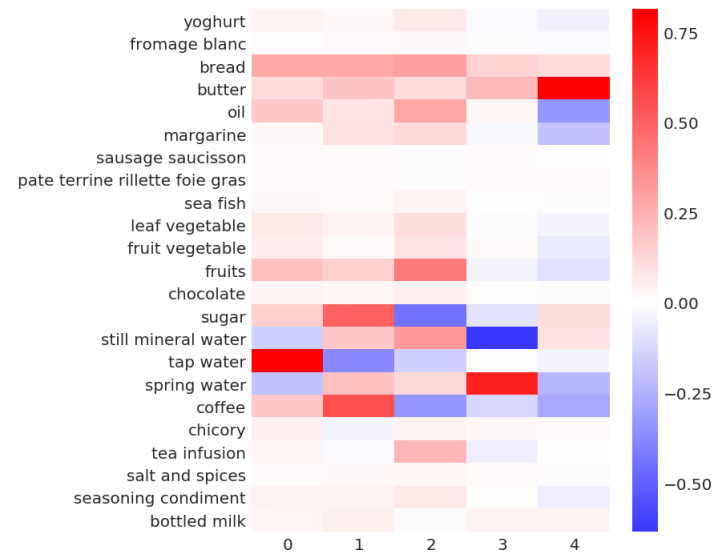


Figure 4: Factor loadings of PCA: explaining the new representation space

### 3.3 Doc2Vec on users

We constitute the corpus by aggregating the food item consumptions per user, each user constituting a document. We use the Gensim implementation of Doc2Vec in order to learn our model. The corpus contains 2624 documents. After learning the model, we compute the cosinus similarity of users and perform spectral clustering. The optimal number of clusters is 5 clusters obtained using the eigengap heuristic.

### 3.4 Doc2Vec on meals

We gather the corpus of meals by aggregating food items consumed at the same moment of consumption, at the same day, by the same user. The corpus is constituted of 37 283 unique meals. A meal embedding is learned using the Gensim Doc2Vec implementation. For each user, the vector of each of his meals is computed leading to user matrices. The similarity matrix between users is obtained by applying the cosine kernel to user matrices. Spectral clustering is applied and the number of clusters is determined using the eigengap heuristic. It yields 3 clusters.

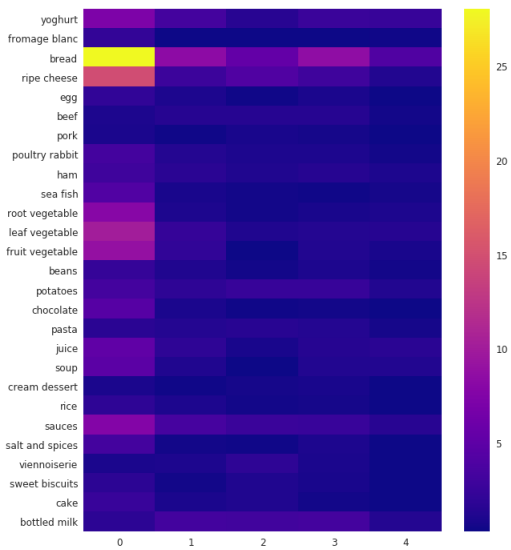


Figure 5: Factor loadings of NMF: explaining the new representation space

### 3.5 Comparison of the clustering results

Our goal now is to compare the clustering results and determine in which cases a food-based approach is adequate and the contribution of a meal-based approach. In order to compare agreement between clustering results, we compute the Adjusted Rand Index (ARI) [16]. It is a popular measure which consists in computing the agreement between two clustering results i.e two partitions. ARI is recommended for cases where the number of clusters is different, which is our case. ARI takes values in  $[-1, 1]$ , 1 meaning that both clusterings agree, values close 0 mean that clusterings are made at random.

	FOOD BASED			MEAL BASED
	PCA	NMF	Doc2Vec users	Doc2Vec meals
PCA	1	0,93	0,14	0,017
NMF		1	0,13	0,018
Doc2Vec users			1	0,013
Doc2Vec meals				1

Table 2: Comparison of clustering results with Adjusted Rand Index

We also plot in Figure 6 the repartition of users in clusters across the methods. From one method to another, the number of cluster is attributed randomly and does not hold meaning.

## 4 DISCUSSION

### 4.1 Comparison of PCA and NMF for food based user modelling

No matter the factorization method used before the clustering step, the clustering results are very similar according to the Adjusted Rand index. This means that the choice of the factorization method

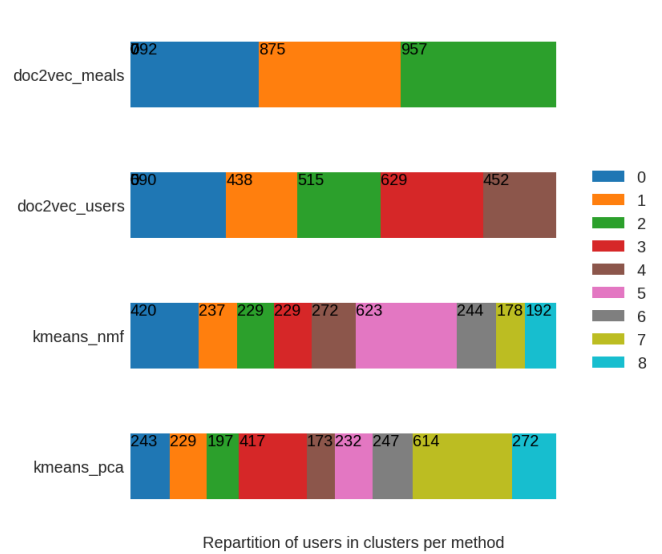


Figure 6: Repartition of users in clusters per method

for clustering users based on their food consumptions is not primordial. However, as shown in Figures 4 and 5, the eating behaviours discovered are different. The coefficients of PCA can be interpreted as consumptions when positive and non consumption when negative. For instance, the eating behaviour 0 consists in drinking tap water but not spring or mineral water. We can also extract information such that those who consume coffee do not consume tea and vice versa. On the opposite, the coefficients of NMF are strictly positive hence the interpretation only concerns food consumptions. For instance, the eating behaviour 0 consists in eating all types of vegetables. The extracted eating behaviours are different according to the method of reduction of dimensionality. We recommend to test both methods to compare extracted eating behaviours as the provided insights of both methods can be interesting.

### 4.2 Contribution of Doc2Vec for the food based approach

We apply Doc2Vec directly to users in order to challenge the state of the art methods in food based approaches as we want to see how the NLP method performs on this task. The number of clusters using the Doc2Vec method on users yields a smaller number of clusters and clustering results are rather different. A major drawback of this method is that eating behaviours cannot be inspected as easily as in the state of the art methods. Further analysis is needed in order to understand why clustering results are so different. This method is adequate if the objective is to extract clusters of consumers, however in this state, this approach is not really adapted if explanations are expected about eating behaviours. Being able to identify eating behaviours is key for recommendation purposes as explanations may be needed for people to implement the recommendations. Usually, the performance of a neural language model is computed on supervised tasks such as document retrieval or analogies. We are in an unsupervised setting which complicates the assessment of the performance of the learned meal embedding.

### 4.3 Comparison state in the art food based approach and meal based approach

It is in the meal based approach that the number of clusters is the smallest. This shows that consumers of this dataset with regards to their way of composing their meals are less diverse as we only find 3 clusters. This result should be interpreted in the light of the assumption made about eating behaviours. We consider that two consumers are similar in the meal based approach if they consume similar meals on the same moment of the day on the same day, a strong assumption on 7-day food diary data. This may lead to more or less low values of similarity overall between users yielding in lesser clusters. It would be interesting to investigate the relaxation of this assumption by assuming that users are similar if they consume similar meals regardless the day of consumption or the moment of consumption. Again, it is difficult to extract eating behaviours as the model is not designed for this purpose. Another language model could be used for modelling food consumption, Latent Dirichlet Allocation (LDA) model.

## 5 CONCLUSION

In this paper we explore user modelling in food consumption for clustering users for recommendation purposes. We compare two state of the art methods in the nutrition community. Our conclusion is that both methods yield more or less the same clustering results. However, the eating behaviours discovered are different. Moreover, we propose a new food-based approach by considering food consumptions as textual data and learning an embedding model with Doc2Vec. The application of Doc2Vec to user food consumption is adequate for user clustering, however it is not adapted for extracting eating behaviours. We argued the importance of having a holistic approach toward nutrition in order to make acceptable recommendations. We propose a new meal based approach which consists in learning a meal embedding space and then computing user similarity based on their meals' similarity. The usage of NLP for food data analysis is promising. However, if clusters that can be explained is needed (which is often the case), then it is better to resort to generative language models such as LDA. Further work will investigate the use of LDA for modelling eating behaviours.

## 6 ACKNOWLEDGEMENT

This study was funded by Danone Nutricia Research.

## REFERENCES

- [1] BIER, D. M., DERELIAN, D., GERMAN, J. B., KATZ, D. L., PATE, R. R., AND THOMPSON, K. M. Improving compliance with dietary recommendations. *Nutrition Today* 43, 5 (sep 2008), 180–187.
- [2] DUBUISSON, C., LIORET, S., TOUVIER, M., DUFOUR, A., CALAMASSI-TRAN, G., VOLATIER, J.-L., AND LAFAY, L. Trends in food and nutritional intakes of french adults from 1999 to 2007: results from the INCA surveys. *British Journal of Nutrition* 103, 07 (dec 2009), 1035.
- [3] ELSWEILER, D., AND HARVEY, M. Towards automatic meal plan recommendations for balanced nutrition. In *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16-20, 2015* (2015), pp. 313–316.
- [4] FREYNE, J., AND BERKOVSKY, S. Recommending food: Reasoning on recipes and ingredients. In *User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings* (2010), pp. 381–386.
- [5] GE, M., ELAHI, M., FERNAÁNDEZ-TOBIÁS, I., RICCI, F., AND MASSIMO, D. Using tags and latent factors in a food recommender system. In *Proceedings of the 5th International Conference on Digital Health 2015* (New York, NY, USA, 2015), DH '15, ACM, pp. 105–112.
- [6] HAGEMAN, P. A., PULLEN, C. H., HERTZOG, M., AND BOECKNER, L. S. Effectiveness of tailored lifestyle interventions, using web-based and print-mail, for reducing blood pressure among rural women with prehypertension: main results of the wellness for women: DASHing towards healthclinical trial. *International Journal of Behavioral Nutrition and Physical Activity* 11, 1 (dec 2014).
- [7] HOFFMANN, I. Transcending reductionism in nutrition research. *The American Journal of Clinical Nutrition* 78, 3 (sep 2003), 514S–516S.
- [8] IVENS, B. J., AND SMITH EDGE, M. Translating the Dietary Guidelines to Promote Behavior Change: Perspectives from the Food and Nutrition Science Solutions Joint Task Force. *J Acad Nutr Diet* 116, 10 (Oct 2016), 1697–1702.
- [9] LAU, J. H., AND BALDWIN, T. An empirical evaluation of doc2vec with practical insights into document embedding generation. In *Proceedings of the 1st Workshop on Representation Learning for NLP, RepNLP@ACL 2016, Berlin, Germany, August 11, 2016* (2016), pp. 78–86.
- [10] LE, Q., AND MIKOLOV, T. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32* (2014), ICML'14, JMLR.org, pp. II–1188–II–1196.
- [11] LEE, D. D., AND SEUNG, H. S. Learning the parts of objects by nonnegative matrix factorization. *Nature* 401 (1999), 788–791.
- [12] LUO, X., ZHOU, M., XIA, Y., AND ZHU, Q. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Trans. Industrial Informatics* 10, 2 (2014), 1273–1284.
- [13] LUXBURG, U. A tutorial on spectral clustering. *Statistics and Computing* 17, 4 (Dec. 2007), 395–416.
- [14] MIJANGOS, V., SIERRA, G., AND MONTES, A. Sentence level matrix representation for document spectral clustering. *Pattern Recognition Letters* 85 (2017), 29–34.
- [15] NEWBY, P. K., AND TUCKER, K. L. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr. Rev.* 62, 5 (May 2004), 177–203.
- [16] RAND, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66, 336 (dec 1971), 846–850.
- [17] REEDY, J., WIRFALT, E., FLOOD, A., MITROU, P. N., KREBS-SMITH, S. M., KIPNIS, V., MIDTHUNE, D., LEITZMANN, M., HOLLENBECK, A., SCHATZKIN, A., AND SUBAR, A. F. Comparing 3 dietary pattern methods—cluster analysis, factor analysis, and index analysis—with colorectal cancer risk: The NIH-AARP diet and health study. *American Journal of Epidemiology* 171, 4 (dec 2009), 479–487.
- [18] ROUSSEUW, P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 1 (Nov. 1987), 53–65.
- [19] SHIM, J.-S., OH, K., AND KIM, H. C. Dietary assessment methods in epidemiologic studies. *Epidemiology and Health* (jul 2014), e2014009.
- [20] TENG, C.-Y., LIN, Y.-R., AND ADAMIC, L. A. Recipe recommendation using ingredient networks. In *Proceedings of the 4th Annual ACM Web Science Conference* (New York, NY, USA, 2012), WebSci '12, ACM, pp. 298–307.
- [21] THORPE, M. G., MILTE, C. M., CRAWFORD, D., AND MCNAUGHTON, S. A. A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older australians. *International Journal of Behavioral Nutrition and Physical Activity* 13, 1 (feb 2016).
- [22] WEBB, D., AND BYRD-BREDBENNER, C. Overcoming consumer inertia to dietary guidance. *Advances in Nutrition* 6, 4 (jul 2015), 391–396.
- [23] WENDEL, S., DELLAERT, B. G., RONTETAP, A., AND VAN TRIJP, H. C. Consumers' intention to use health recommendation systems to receive personalized nutrition advice. *BMC Health Services Research* 13, 1 (apr 2013).
- [24] WOOLHEAD, C., GIBNEY, M. J., WALSH, M. C., BRENNAN, L., AND GIBNEY, E. R. A generic coding approach for the examination of meal patterns. *The American Journal of Clinical Nutrition* 102, 2 (jun 2015), 316–323.
- [25] WORLD HEALTH ORGANIZATION. Diet, nutrition and the prevention of chronic diseases: report of a joint who/fao expert consultation. Tech. rep., 2003.
- [26] ZETLAOU, M., FEINBERG, M., VERGER, P., AND CLEMENÇON, S. Extraction of food consumption systems by nonnegative matrix factorization (NMF) for the assessment of food choices. *Biometrics* 67, 4 (2011), 1647–1658.