Development of a machine learning-based models for recommending substitutable ingredients

Ania Wróblewska

Faculty of Mathematics and Information Science
Warsaw University of Technology

Project TAISTI: Artificial Intelligence for Inferring Ingredient Substitutes in Culinary Recipes
Kierownik: Agnieszka Ławrynowicz, Poznan University of Technology

Norway grants

The National Centre for Research and Development

Our goal

- Recommendations
- Recommendations for food substitutes in recipe datasets

Recommendation systems - automatic filtering techniques to facilitate a user search

Recommenders' goals:

- reducing the amount of data
- selecting the more relevant data for the user

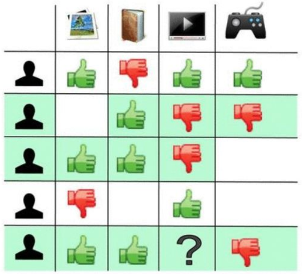
Goal

Personalized information retrieval based on:

- user preferences and their historical data analysis,
- the whole community analysis and needs (from the point of view of users and item vendors),
- the recommended items' characteristics.

Recommenders

- Information Retrieval (Search Engines) vs. Information Filtering (Recommender Systems)
  - Query information retrieving vs. Query-less search



- Recommendation methods
  - Popularity-Based, Collaborative Filtering (item/user-based)
  - Content-Based
  - Hybrid

- Evaluation & Metrics:
  - MAP@K, MAR@K (recall at the kth recommendations), Catalogue Items Coverage

* https://github.com/microsoft/recommenders

https://github.com/statisticianinstilettos/recmetrics

https://github.com/dg4271/Deep-Learning-for-Recommendation-System

Baselines

Librec - java library

https://guoguibing.github.io/librec/index.html

RecBole - python library

https://recbole.io/

Datasets:

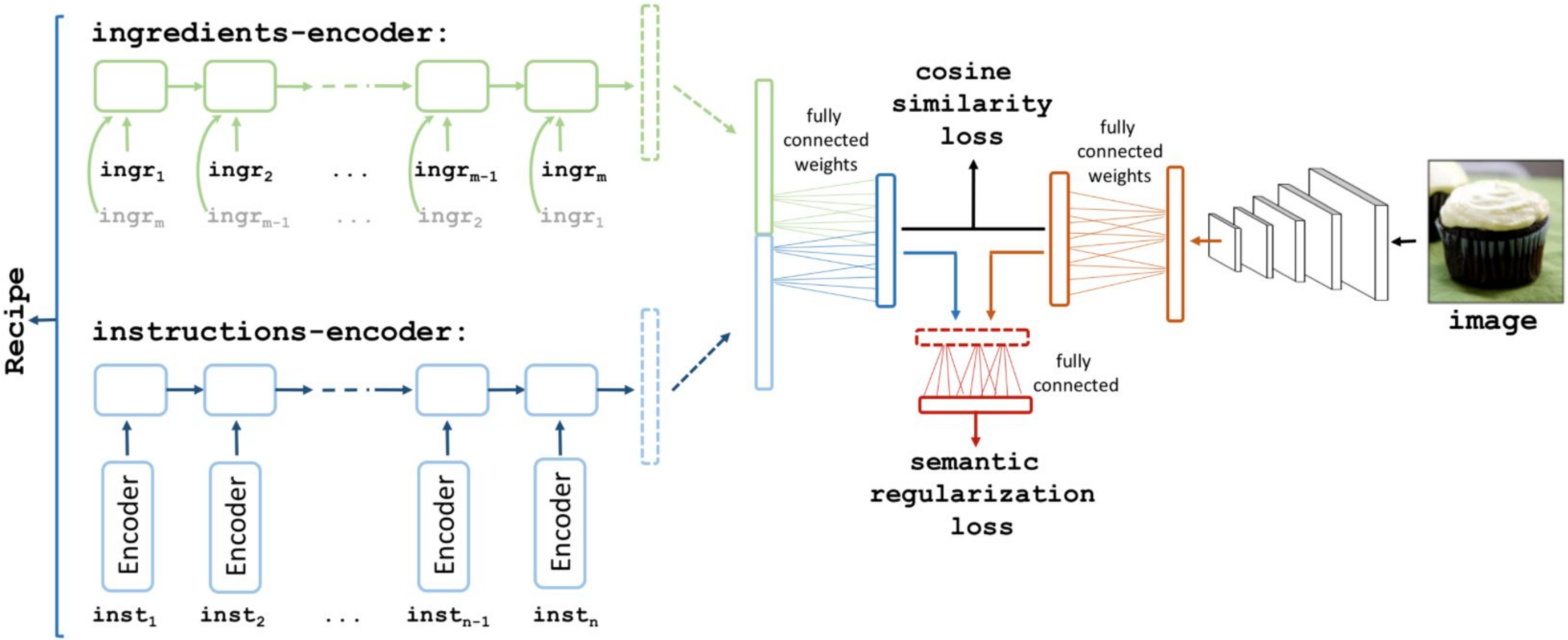
https://github.com/RUCAIBox/RecSysDatasets

Embeddings

- Vector representations for TEXT and other modalities
  - Bag-of-words, one-hot, TF-IDF…
  - Modern: Neural networks, word2vec, GloVe ~ co-occurence matrix, FastText, n-gram/subwords, deep contextualized models
  - Other types of data, uni/multimodal data

- Evaluation:
  - INTRINSIC - direct (language modeling, word relations: similarity, analogy)
  - EXTRINSIC - indirect (using in other tasks, e.g. POS, sentiment analysis)

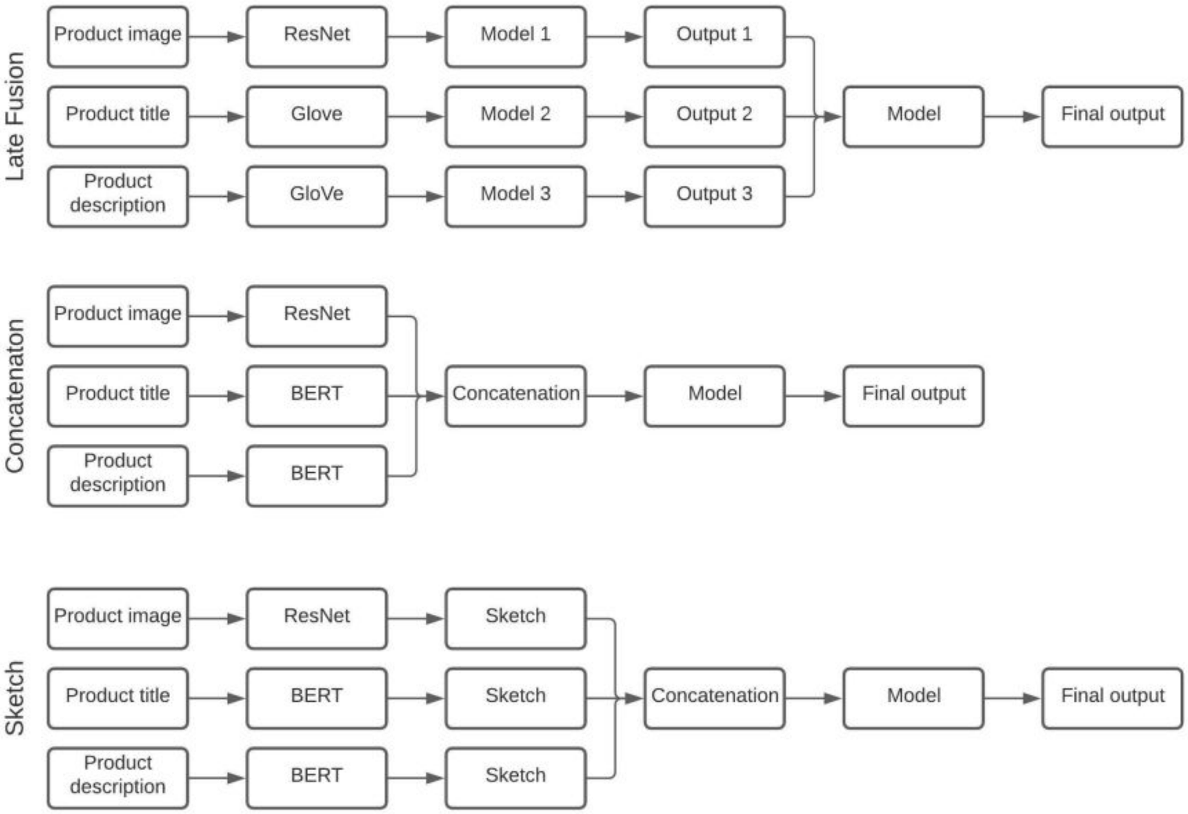
Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C.J. (2019). Evaluating word embedding models: methods and experimental results. *APSIPA Transactions on Signal and Information Processing, 8*. https://doi.org/10.1017/atsip.2019.12 https://arxiv.org/abs/1901.09785



https://deepai.org/publication/recipe1m-a-dataset-for-learning-cross-modal-embeddings-for-cooking-recipes-and-food-images

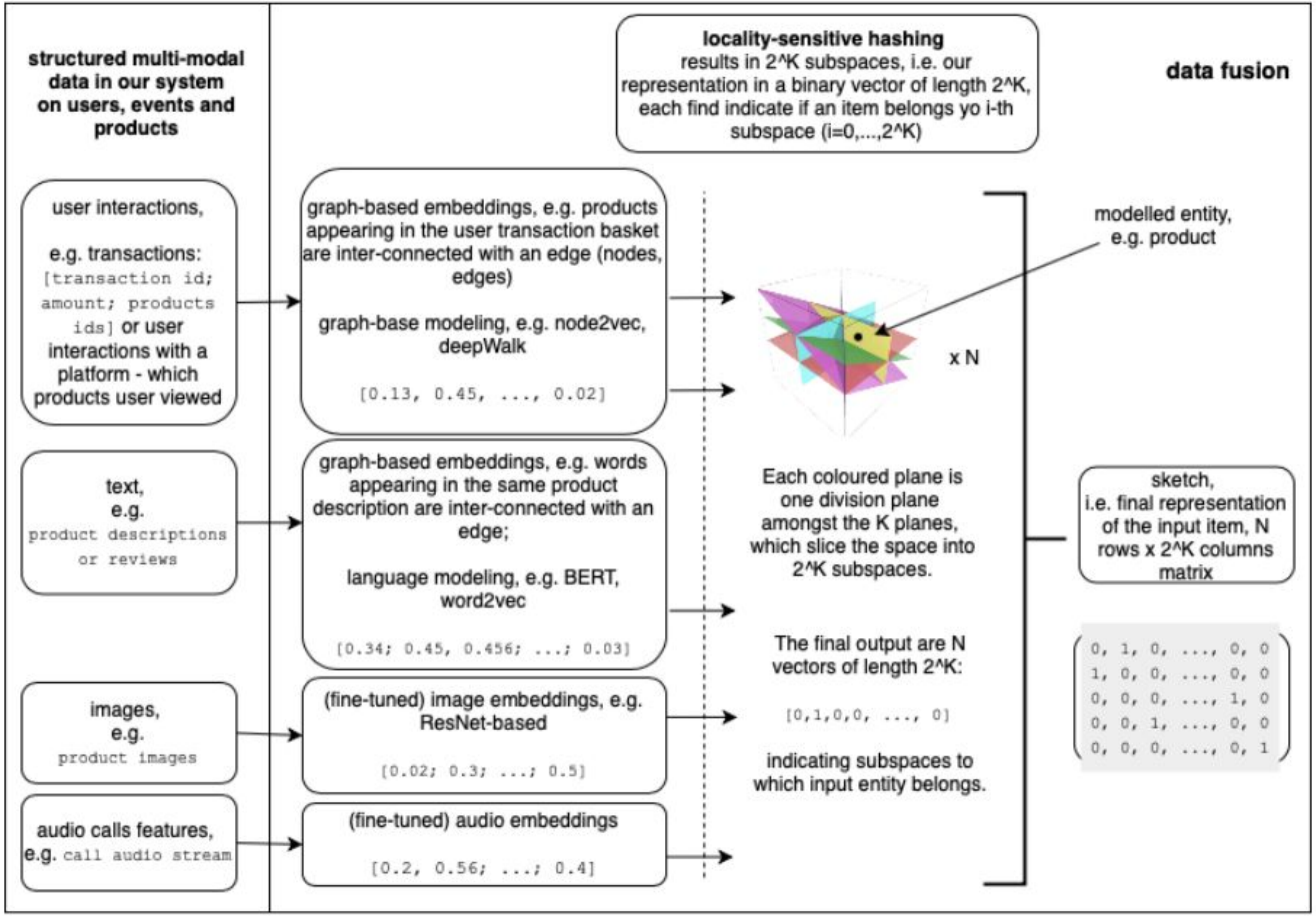
Marín, Javier & Biswas, Aritro & Ofli, Ferda & Hynes, Nicholas & Salvador, Amaia & Aytar, Yusuf & Weber, Ingmar & Torralba, Antonio. (2018). Recipe1M: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images.

Data Fusion



* Pawłowski, M.; Wróblewska, A.; Sysko-Romańczuk, S. Effective Techniques for Multimodal Data Fusion: A Comparative Analysis. *Sensors* **2023**, *23*, 2381. https://doi.org/10.3390/s23052381

Sketch representation



* Wróblewska, A.; Dąbrowski, J.; Pastuszak, M.; Michałowski, A.; Daniluk, M.; Rychalska, B.; Wieczorek, M.; Sysko-Romańczuk, S. Designing Multi-Modal Embedding Fusion-Based Recommender. *Electronics* **2022**, *11*, 1391. https://doi.org/10.3390/electronics11091391

Goal

Personalized information retrieval based on:

- user preferences and their historical data analysis,
- the whole community analysis and needs (from the point of view of users and item vendors),
- the recommended items' characteristics.

Recipe or ingredient recommendation systems assist users in finding a personalized and balanced diet, encouraging healthy eating habits.

A nourishing diet is critical in maintaining a person's health, yet numerous factors influence people, and therefore it is often challenging to compose healthy recipes and diet.

* Weiqing Min et al. “A Survey on Food Computing”. ACM Comput. Surv. 52.5 (Sept. 2019). https://doi.org/10.1145/3329168
* Tian Y, Zhang C, Metoyer R and Chawla NV (2022) Recipe Recommendation With Hierarchical Graph Attention Network. Front. Big Data 4:778417. doi: 10.3389/fdata.2021.778417

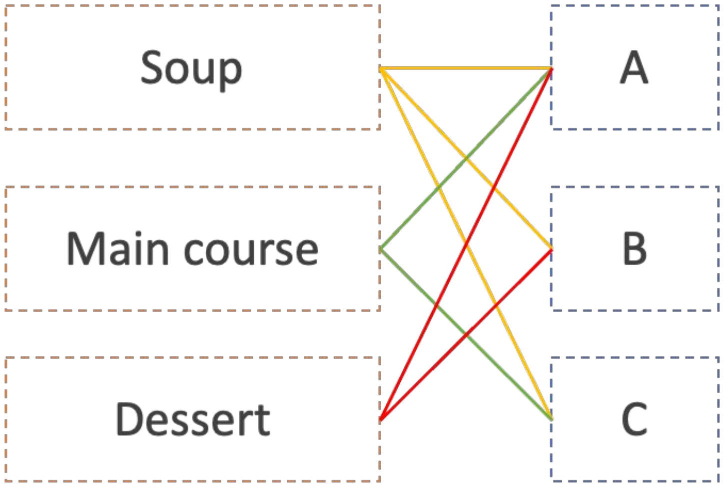
Data

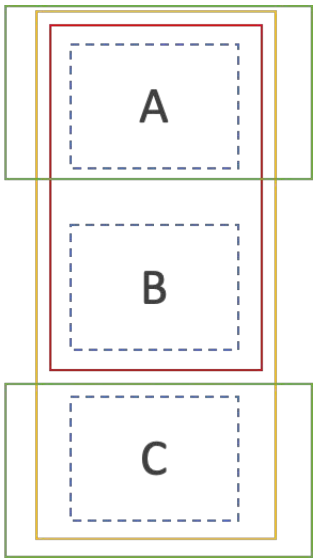
- Behavioural data - collaborative filtering and content-based recommendations
- Metadata, here:
  - recipe texts (ingredient sets and instructions),
  - images of the dishes ?
- Knowledge-based systems also consider additional sources of data, e.g. nutritional ingredient values, food ontologies, thesauri, and context-based systems regarding context, e.g. health needs and individual preferences

Datasets:
https://github.com/RUCAIBox/RecSysDatasets

Representation learning - hypergraph model

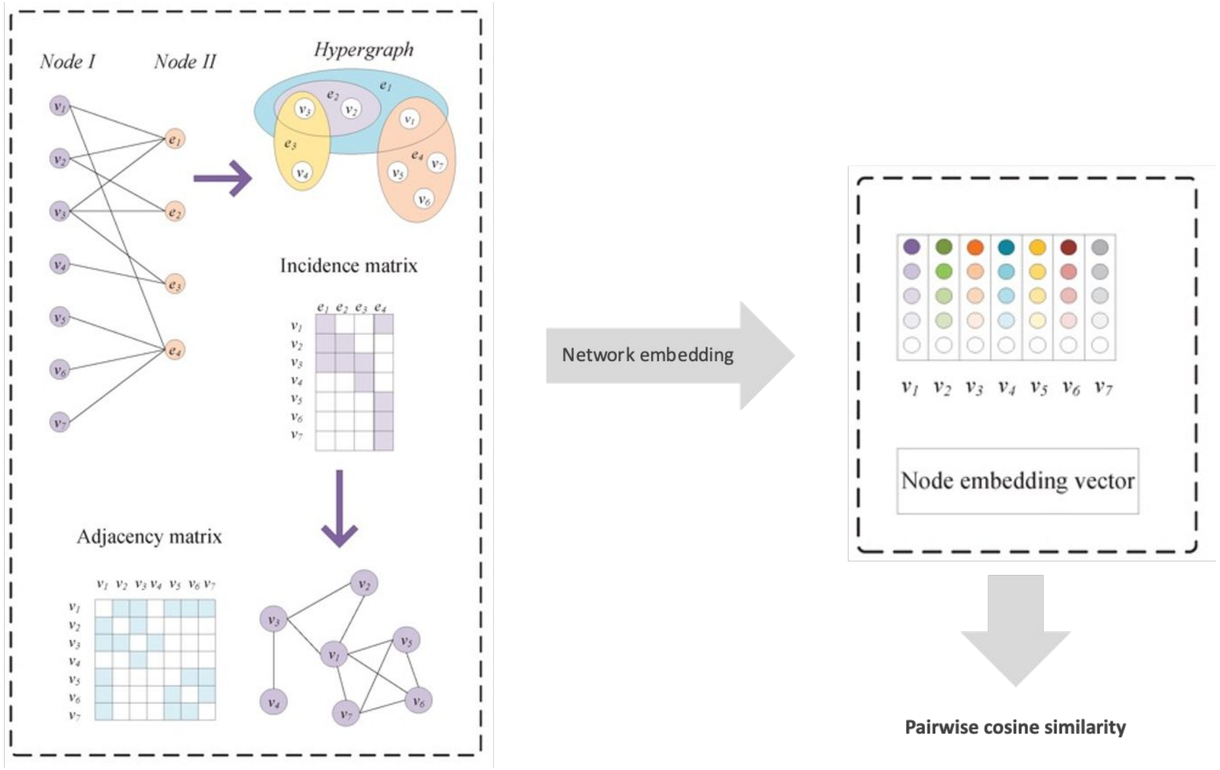
Recipe	Ingredient
Soup	A
Soup	B
Soup	C
Main course	A
Main course	C
Dessert	B
Dessert	A



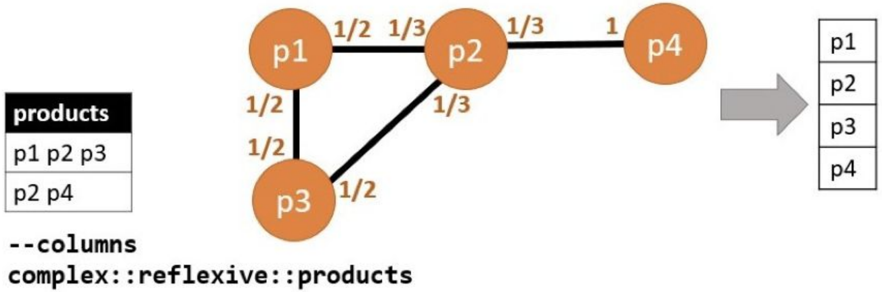


Norway
grants

Representation learning - hypergraph model



Recommenders



"lemon": ["orange juiced", "freshsqueezed lemon juice", "chavrie goat cheese", "blueberry vinegar"]

"chocolate fudge cake": ["powdered sugar", "stiff whipping cream", "cinnamon dolce", "chocolate butter", "bakers white chocolate"]

"frozen strawberries": ["crushed apricots", "strawberry jello", "frozen strawberries with sugar", "flavor gelatin", "glass apricot"]

Norway grants

* B. Rychalska et al: “Cleora: Graph embeddings” https://arxiv.org/pdf/2102.02302.pdf

* Sergiy Tkachuk, Anna Wróblewska, Jacek Dąbrowski, Szymon Łukasik, “Identifying Substitute and Complementary Products for Assortment Optimization with Cleora Embeddings” https://arxiv.org/abs/2208.06262

Our baselines

PPMI to model the probabilities that:

- recipe and ingredient occur together
- two different ingredients occur together

$$PPMI(x;y) \equiv \max\left(\log_2 \frac{p(x,y)}{p(x)p(y)}, 0\right)$$

FastText algorithm utilizing ingredient entities
Frequent Sets algorithms

Ingredient being substituted	PPMI recipe oriented	PPMI word oriented	FastText recipe oriented	FastText ingredient oriented	FreqSets
chicken	breast	chicken broth	chicken wing	chicken part	carrot, tomato, oliv oil, parsley, beef
tofu	sesam oil	tamari	bean sprout	sesam oil	not present
milk	biscuit mix	egg	skim milk	carnat milk	flour, bake powder, butter, bake soda, cinnamon
egg	sugar	sugar	egg yolk	egg yolk	butter, soda, bake soda, nut, cinnamon
banana	banana extract	vanilla wafer	peach	angel flake coconut	vanilla, bake powder, butter, cinnamon, bake soda
butter	sugar	margarin	margarin	margarin	salt, bake powder, soda, bake soda, milk

Norway grants

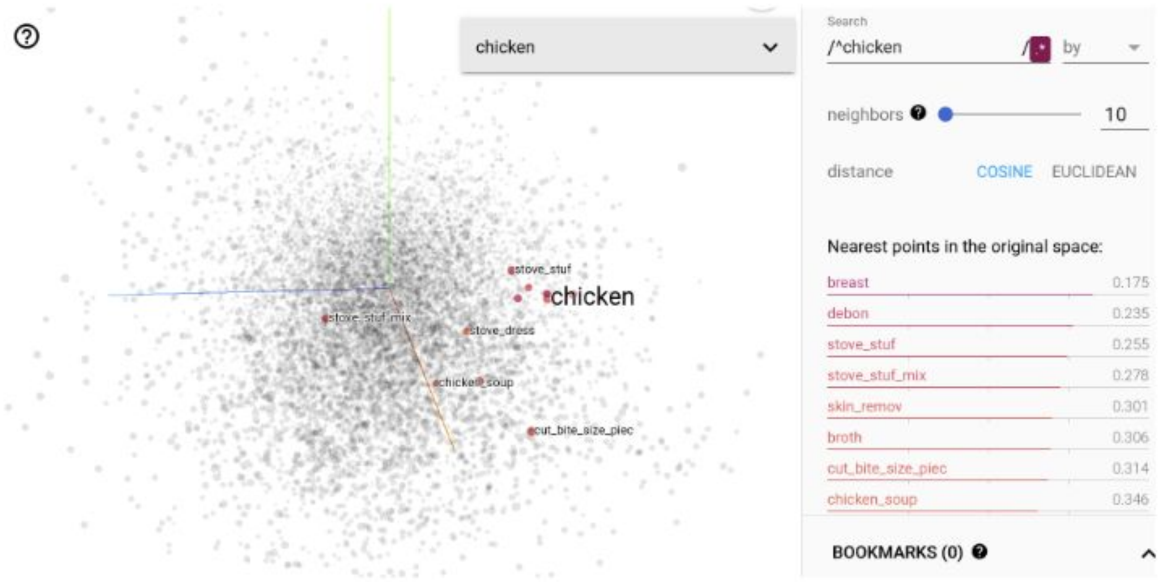


Figure 19: Visualisation of PPMI recipe oriented for chicken *RecipeNLG*

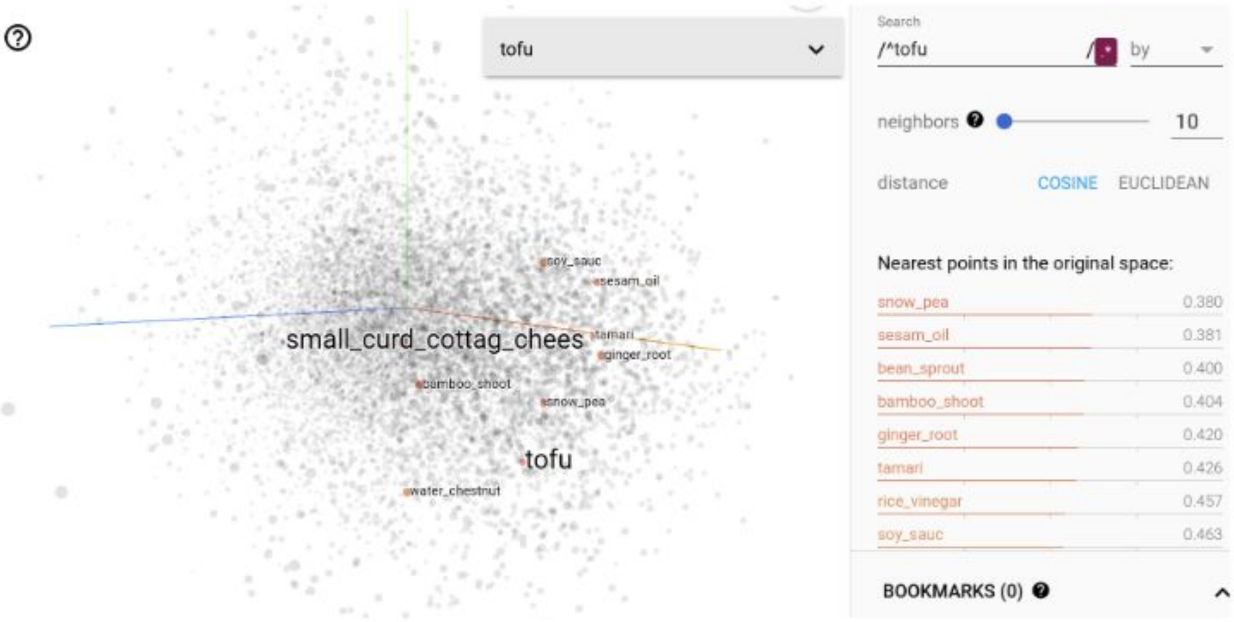


Figure 21: Visualisation of PPMI recipe oriented for tofu *RecipeNLG*

Conclusions

- We implemented substitutes' recommendation baselines comprising modification of PPMI, FastText algorithms, and methods for calculating frequent itemsets.
- No benchmark datasets with gold standards to evaluate the approach. However, we utilized visualisations and listing to be further assessed by experts in dietary and food technology.

More important issues related to our task are:

- FOOD entities are not grouped, which led to less reliable results, e.g. ”egg” is treated as completely different as ”egg yolk”. To overcome this, we need a limited vocabulary to assign each FOOD entities into similar subgroups.
- The automatic recommendation results are sometimes not usable; how-ever, these results should be taken after imposing restrictions on dietary needs or functional attributes of the food ingredients.

Norway grants