Bargaining agents based system for automatic classification of potential allergens in recipes

José Alemany, Stella Heras, Javier Palanca, and Vicente Julián

Departamento de Sistemas Informaticos y Computacion, Universitat Politècnica de València (Spain)
{jalemany1, sheras, jpalanca, vinglada}@dsic.upv.es

KEYWORD

Recommendation system; Food allergy; Multi-agent system

ABSTRACT

The automatic recipe recommendation which take into account the dietary restrictions of users (such as allergies or intolerances) is a complex and open problem. Some of the limitations of the problem is the lack of food databases correctly labeled with its potential allergens and non-unification of this information by companies in the food sector. In the absence of an appropriate solution, people affected by food restrictions cannot use recommender systems, because this recommend them inappropriate recipes. In order to resolve this situation, in this article we propose a solution based on a collaborative multi-agent system*, using negotiation and machine learning techniques, is able to detect and label potential allergens in recipes. The proposed system is being employed in* `receteame.com`*, a recipe recommendation system which includes persuasive technologies, which are interactive technologies aimed at changing users' attitudes or behaviors through persuasion and social influence, and social information to improve the recommendations.*

1. Motivation

When the doctor diagnoses a patient with an allergy, the set of substances that produce extreme sensitivity in the patient's immune system should be avoided. In the case of food allergies, the difficulty is greater because avoiding a food is not an easy task when many of them are composed by others (e.g. mayonnaise, is composed of oil, egg, vinegar, etc.). This is an awkward situation of high impact because it involves nutrition, a necessary and daily task in the lives of people activity. This increasingly affects to more individuals in our society (up to 8% in children and 2% in adults)[1]. Moreover, the problem of food allergies is not resolved by simply avoiding certain foods, because the lack of nutrients they provide must be compensated with other foods.

People affected by food allergies are forced to become expert nutritionists to maintain a healthy life, free from allergens that they cannot tolerate. Currently, Internet is the most popular way of obtaining information about allergies. On the Internet, for instance, the World Allergy Organization (WAO) regulates and offers the terminology used to characterize allergies information. However, the information is difficult to understand because of its complexity and quantity. Thus, traditional search and navigation activities are being combined or even replaced by direct interactions between users in the form of recommendations, advice and warnings; 2 out of 3 take into account the recommendations of other users to make decisions (about products, treatments, entertainment, etc.); and of these, 69% gives a lot or some credibility to what their friends or acquaintances say on social networks.

[1] WAO World Allergy Organization, Food allergy statistics: http://www.worldallergy.org/public/allergic_diseases_center/foodallergy/

However, the growing number of users and information generated, the heterogeneity of users, their unpredictable behavior, and the dynamism of the structure of social networks cause users a high degree of uncertainty when choosing with whom to interact and what information they should consume (Van der Aalst and Song, 2004). To reduce this uncertainty, tools that help users in their decision-making processes are required. A promising solution is the use of recommendation systems (Adomavicius and Tuzhilin, 2005; Zhou et al., 2012), which are able of performing effective recommendations to help users to make appropriate decisions.

In (Palanca et al., 2014), we presented `receteame.com`, a persuasive social recommendation system whose goal is to recommend the most appropriated recipe to each specific user, taking into account their preferences, food restrictions and social context. In this work, we propose an improvement of the system to automatically detect allergens in recipes, based on their ingredients composition, and to prevent the system from recommending inappropriate recipes to people with specific food restrictions. `receteame.com` uses the USDA[2] nutrient database to provide more detailed nutritional information about the ingredients of the recipes. However, this database does not include full information about the potential allergies associated with each ingredient and, to the best of our knowledge, there is no reference database that labels ingredients with their associated allergens.

This is a major problem for our system, since an accurate and reliable detection of recipe allergens must be performed each time a new recipe is uploaded or updated in the system, and always before the system can provide a specific recommendation. Thus, the system must avoid errors in the classification of allergens in a recipe, as this may drastically decrease its reputation among users affected by food allergies (a population that is particularly reluctant to rely on information from the Web or computer systems) and even, in the worst case, could cause serious health problems to their users. The solution that we propose in this paper consists on a collaborative *multi-agent expert system* that is able to automatically detect food allergies in nutrients and label ingredients with their potential allergens.

2. Related work

As the prevalence of diet-related diseases (either those directly caused by poor nutrition or those that condition the diet of their affected patients) is growing, the interest of the artificial intelligence community to develop intelligent systems to help humans with their food needs is also on the rise.

Early research on planning meals was started in the area of case-based reasoning, where the system JULIA (Kolodner, 1987) created plan meals and the system CHEF (Hammond, 1989) created new recipes based on those it already knew about. More recently, the recommender systems community has also paid attention to this topic, and we can find in the literature some recipe recommendation systems that focus on specific diseases. The food recommendation system presented in (Phanich et al., 2010), which uses food clustering analysis for diabetic patients to recommend the proper substituted foods in the context of nutrition and food characteristic, is an example of this type. Others try to predict recipe ratings, as the system proposed in (Teng et al., 2012), which illustrates how these ratings can be predicted with features derived from combinations of ingredient networks and nutrition information.

Even when people do not have specific food restrictions, the improvement of their diet and health is a hot topic. With this aim, the recipe recommendation system proposed in (Freyne and Berkovsky, 2010) makes tailored recommendations of healthy recipes; in (Ueda et al., 2011) authors created a personalized recipe recommendation method that is based on the user's food preferences from his/her recipe browsing and cooking history; in (Harvey et al., 2013) authors presented a system that learns user tastes for improving the rating prediction and make better recommendations of recipes; and in (Elsweiler and Harvey, 2015) the system recommends recipes that users will like and that fit into a balanced diet.

[2]USDA National Nutrient Database for Standard Reference: http://ndb.nal.usda.gov/

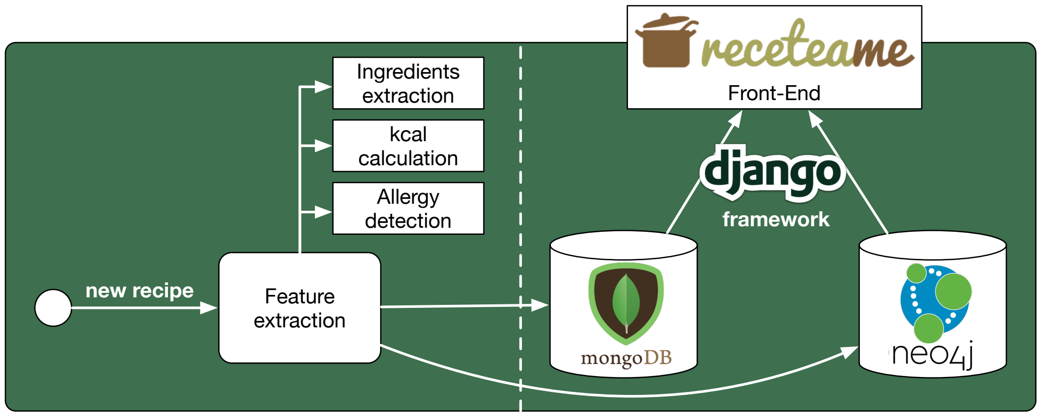

Figure 1: Receteame.com platform: components and services.

In the area of persuasion technologies, the *Portia* system (Mazzotta et al., 2007) used relevant arguments, both rational and emotional, to persuade people to change their eating habits.

However these proposals do not follow a social network-based recommendation approach (Schall, 2015) and are not able to automatically detect and classify allergens in the recipes ingredients. The approach proposed in this work deals with the latter challenge and, to the best of our knowledge, is the first attempt in a recommendation system to include this interesting functionality.

3. Application Design

`receteame.com` is a persuasive social recommendation system that has been designed following the MVC (Model-View-Controller) model. The design of the application is shown in Figure 1. The database contains information about the model. This is the information about our users (personal information, their tastes, preferences, allergies, etc.), and the recipes and ingredients (calories and nutritional information, allergens and others) that we recommend. Moreover, the application includes a front-end (web interface) that allows users to interact with the recommender, and ultimately, with the application. The front-end represents the view and the controller.

Our system can register new recipes in two ways: via other web sites or through new registered entries by users. Each time a new recipe is uploaded in the system, a process for the extraction of characteristics thereof is performed. This process extracts the ingredients of the recipe, detects potential allergens and calculates the kilocalories of the recipe. We use a heuristic to match the ingredients of our recipes with those corresponding ingredients of the USDA nutrient database, which contains detailed data of 8789 food items.

In order to allow the system to detect and classify ingredients according to their potential food allergies, we have designed a collaborative multi-expert-system. Figure 2 shows its design. In the figure, each expert represents a machine learning technique that we have trained to label the ingredients of our recipes with their associated allergens. When a new ingredient is extracted from a new recipe, each expert analyses the ingredient separately. After that, the experts inform to the decision-maker agent whether or not the ingredient has potential food allergens. The decision-maker, according to the opinion of each expert, is then in charge of labeling the ingredient

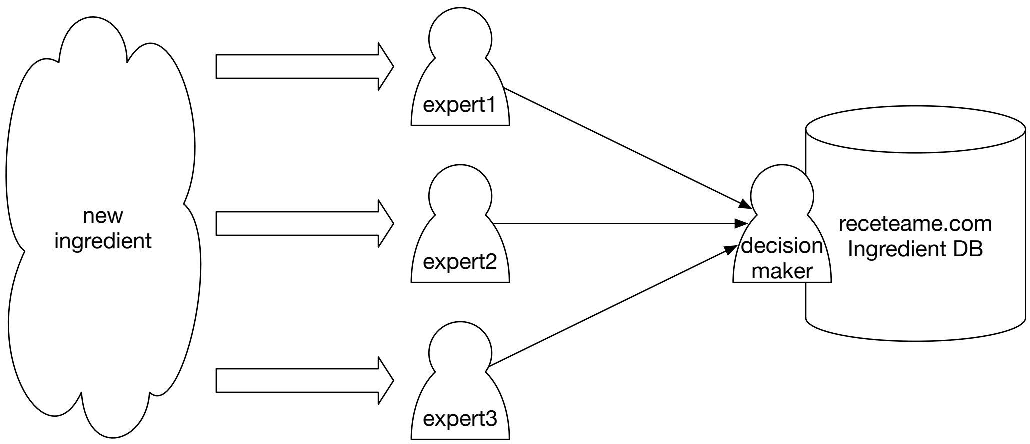

Figure 2: Collaborative multi-expert-system.

with the set of detected allergens. We have implemented three experts for our collaborative multi-expert-system, each representing one of the best machine learning techniques (decision trees, logistic regression and K nearest neighbors) that we have tested for our system.

Since an ingredient may contain more than one allergen, our experts are binary classifiers that the system calls for each specific allergen that we want to detect. As will be explained in the next section, the use of separated classifiers allows us to improve the detection of allergens and also to increase the accuracy and reliability of our system, relying in the vote of the majority to discard individual classification errors in such a sensible domain.

4. Evaluation

As has already been mentioned, the persuasive social recommendation system `receteame.com` uses the USDA nutrient database to provide more detailed nutritional information about recipes. The USDA national nutrient database is an open database produced by the US Department of Agriculture that provides nutritional data of generic and proprietary-branded foods. It is the major source of information with data about 8789 food items and up to 150 food components per item. Most of these values have been calculated analytically or derived from analytical processes by scientists. However, there are still many entries with unknown values in food components. The USDA nutrient database is updated regularly and may be searched or downloaded through a REST API. In Figure 3 you can see an example of one entry of the database.

With the nutritional food information acquired by the USDA nutrient database and the tools and algorithms explained in the Section 3, we are going to classify the ingredients (and in consequence the recipes) by their potential allergens.

According to the WAO and AEPNAA[3] (the most important agency in Spain that provides information about food allergies and helps to improve the safety and quality of life of individuals affected by food allergies) among the most common food allergies in the Spanish population are milk, egg, fish, and wheat. Therefore, we have focused mainly on these food allergies. In addition, we have included in our research other common food allergies

[3]Asociación Española de Personas con Alergia a Alimentos y Látex: http://www.aepnaa.org

Statistics Report: 01035, Cheese, provolone

Return to Search Results Basic Report Full Report (All Nutrients) Download (CSV) Print (PDF)

Nutrient values and weights are for edible portion.

Nutrient	Unit	Value (100g)	Data Points	Std. Error	Min	Max	df	LB	UB	# Studies	Source	NDB Ref	Last Modified
Proximates													
Water	g	40.95	21	0.661	--	--	--	--	--	--	Analytical or derived from analytical	--	11/1976
Energy	kcal	351	--	--	--	--	--	--	--	--	Calculated or imputed	--	02/2009
Energy	kJ	1471	--	--	--	--	--	--	--	--	Calculated or imputed	--	02/2009
Protein	g	25.58	16	0.349	--	--	--	--	--	--	Analytical or derived from analytical	--	11/1976
Total lipid (fat)	g	26.62	24	0.341	--	--	--	--	--	--	Analytical or derived from analytical	--	11/1976
Ash	g	4.71	14	0.155	--	--	--	--	--	--	Analytical or	--	11/1976

Figure 3: USDA National Nutrient Database: Example of nutritional information.

like crustacean, mollusk, soy, celery, lupin and sesame. In this work, our aim was the automatic detection of those food allergies in the ingredients of the recipes of our system.

The specific food allergies that have been included in the study presented an additional complication, this has been due to the number of positive samples available. For this reason, in the case of food allergies in which has not been provided sufficient representative samples (celery, lupine, sesame), we have not been able to address the problem of classification and we have been used the dictionary of keywords to label the set of food.

4.1 System Implementation

As shown in Figure 1, we used the `django` framework to implement `receteame.com`. Furthermore, the `mongodb` and `neo4j` technologies have been used to implement the database. The `mongodb` database technology provides quick access and supports high loads of requests. It has been used to store the information used by `receteame.com` to display the information about the recipes. The `neo4j` database provides flexibility and easy queries. Therefore, it has been used to implement the logic of our recommender system.

The implementation of the collaborative multi-agent expert system has been performed by programming a different machine learning technique in each expert. We tested a wide range of techniques, including decision trees, linear regression, logistic regression, support vector machines and k nearest neighbors (see section 4.2). The better classification results were obtained by the following techniques: decision trees, which allows us to build a rule-based model to predict allergies in the ingredients; logistic regression, which is a linear model where the probabilities describing the possible outcomes of a single trial are modeled using a logistic function; and K nearest neighbors, a classifier based on instances which is computed from a simple majority vote of the nearest neighbors of each point, so finally our system includes an expert for each of these three techniques.

To implement the experts, we have used `scikit-learn`[4] and `bigml`[5], which provide simple and efficient tools for data mining and analysis. By using these tools, we have trained a binary classifier, one for each of the food allergies that we want to detect in our ingredients. In addition, to increase the quality of the classification, we have elicited in each agent a dictionary of keywords with basic ingredients that contain the allergies (according to information provided by WAO, AEPNAA, FARE[6] among others). We have included all the models created with

[4]http://scikit-learn.org/
[5]https://bigml.com/
[6]Food Allergy Research & Education: http://www.foodallergy.org

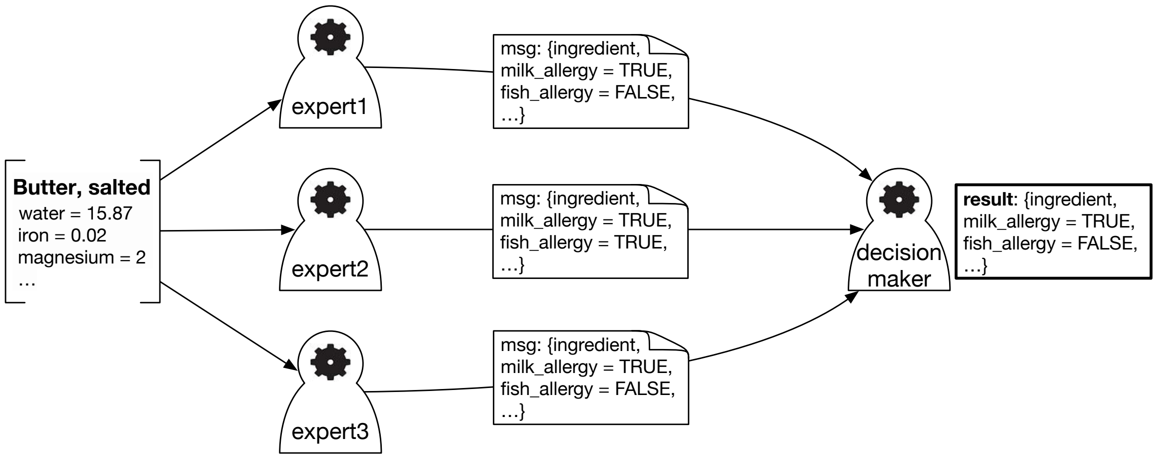

Figure 4: Food allergy detection: Example of collaboration.

each machine learning technique in an individual agent. Therefore, each expert agent has a different model for each allergy, but all allergy models have been trained by using the same machine learning technique in the agent.

The operation of our system is performed as follows: when a new recipe is uploaded or updated in `receteame.com`, the system detects its ingredient composition and sends the list of ingredients to the three expert agents. Then, for allergy to detect, they use the learned models to classify the ingredient (as having or not such allergy) and send a message to the decision-maker agent with the results of the classification. Finally, the decision-maker agent implements a voting system by which the classification agreed by the majority is selected as the final classification for each allergy in each ingredient (see Figure 4 for an example). As will be shown in the next section, with this simple voting technique, our systems reaches good allergy-detection results and is robust against individual misclassification errors.

4.2 Results

To test our system, for each allergy to detect we have created a set of ingredients with and without the allergen. Also, each set has been divided into two subsets: 65% for training and 35% for testing under different machine learning techniques. The results are shown in Table 1[7].

From the results, we can observe how logistic regression, decision-tree and k nearest neighbors techniques achieve the best classification results, so we included these techniques in our expert agents. In addition, Figure 2 shows more detail regarding the percentages of ingredients that have been correctly classified as having the allergen (true positives, labeled as true in the table) or not having the allergen (true negatives, labeled as false in the table).

In our domain, to correctly detect those ingredients that can provoke an allergic reaction in our users is crucial. Therefore, although we achieved quite good classification results with the logistic regression technique alone, we needed to increase the accuracy in detecting true positives and we developed the collaborative multi-expert system presented in section 3. The results obtained by each expert agent alone and by the collaborative system can be seen in Table 2.

[7] We only show in the table a subset of all machine learning techniques that we tested

	Logistic regression (1)	**Decision tree (2)**	**K nearest neighbors (3)**	Support vector machines	Linear regression
Milk allergy	**95.15%**	**89.91%**	**87.67%**	79.66%	74.58%
Egg allergy	**96.50%**	**90.33%**	**86.08%**	80.85%	84.01%
Fish allergy	**98.18%**	**96.04%**	**93.12%**	81.46%	79.93%
Gluten allergy	**96.97%**	**89.01%**	**86.97%**	76.08%	83.61%
Crustacean allergy	**80.24%**	**88.00%**	**86.92%**	54.05%	68.84%
Mollusk allergy	**81.49%**	**79.20%**	**73.99%**	72.93%	75.61%
Soy allergy	**93.50%**	**77.89%**	**81.60%**	73.37%	61.50%

Table 1: Accuracy of the different models applied to sets of each food allergy test.

		Collaborative multi-expert system	Expert1 (1)	Expert2 (2)	Expert3 (3)
Milk allergy	True	**97.72%**	92.91%	87.50%	73.88%
	False	**96.91%**	95.82%	90.38%	91.74%
Egg allergy	True	94.20%	**95.24%**	82.27%	75.68%
	False	95.58%	**96.92%**	93.30%	89.56%
Fish allergy	True	**99.12%**	94.51%	93.85%	73.33%
	False	98.45%	**98.58%**	98.24%	95.28%
Gluten allergy	True	**96.27%**	95.05%	86.47%	76.38%
	False	95.03%	**97.76%**	91.56%	91.29%
Crustacean allergy	True	**100.00%**	61.11%	76.47%	77.78%
	False	99.37%	99.37%	**99.53%**	96.06%
Mollusk allergy	True	**92.86%**	64.71%	59.38%	50.00%
	False	98.46%	98.26%	**99.02%**	97.97%
Soy allergy	True	**97.37%**	88.24%	57.14%	72.73%
	False	98.07%	**98.76%**	98.64%	97.62%

Table 2: Accuracy of collaborative multi-expert system compared to each expert individually (for food allergies treated).

Results demonstrate that the collaborative system, where the classification of the majority is selected as the final classification, improves the general accuracy and more specifically the percentage of true positives (the detection of the ingredients that have the allergen) for the majority of allergies tested. Only in the case of the egg allergy, the logistic regression technique (implemented in expert 1) gets better results alone, but the collaborative technique still achieves similar good results. Therefore, this technique improves the reliability of our system and provides better allergy-detection results.

5. Conclusions and future work

This paper has presented a *multi-agent system* that is able to automatically detect food allergies in nutrients and label ingredients with their potential allergens. The system has been designed as a set of experts, where each expert represents a machine learning technique that tries to label new ingredients. Moreover, the system includes a decision-maker agent which, according to the opinion of each expert, makes a decision about the allergens associated to the new recipes uploaded or updated in the system. At this moment, the number of expert agents implemented is three, but it could be easily extended if other new labeling techniques get better classification results in the future. The proposed system has been integrated into `receteame.com`, a website that uses a persuasive social recommendation system to recommend recipes taking into account the preferences and food restrictions of its users.

Some experiments have been performed in order to validate the proposal. Results show that the combination of separated classifiers into a collaborative technique allows the system to improve the detection of allergens and also to increase the accuracy and reliability, relying in the vote of the majority to discard individual classification errors. In our system, the reliability of the allergens information provided is crucial, since misclassifications could seriously compromise the reputation of the system and in the worst case, could cause serious health problems to our users. Therefore, future work will be performed to investigate new machine learning techniques and voting mechanisms that could decrease classification errors.

Acknowledgements

This work was supported by the projects TIN2015-65515-C4-1-R and TIN2014-55206-R of the Spanish government and by the grant program for the recruitment of doctors for the Spanish system of science and technology (PAID-10-14) of the Universitat Politècnica de València.

6. References

Van der Aalst, W. M. and Song, M., 2004. Mining Social Networks: Uncovering interaction patterns in business processes. In *Business Process Management*, pages 244–260. Springer.

Adomavicius, G. and Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749.

Elsweiler, D. and Harvey, M., 2015. Towards automatic meal plan recommendations for balanced nutrition. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 313–316. ACM.

Freyne, J. and Berkovsky, S., 2010. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 321–324. ACM.

Hammond, K. J., 1989. *Case-based Planning: Viewing Planning As a Memory Task*. Academic Press Professional, Inc., San Diego, CA, USA. ISBN 0-12-322060-2.

Harvey, M., Ludwig, B., and Elsweiler, D., 2013. You are what you eat: Learning user tastes for rating prediction. In *String Processing and Information Retrieval*, pages 153–164. Springer.

Kolodner, J. L., 1987. Capitalizing on failure through case-based inference. Technical report, DTIC Document.

Mazzotta, I., De Rosis, F., and Carofiglio, V., 2007. Portia: A user-adapted persuasion system in the healthy-eating domain. *Intelligent Systems, IEEE*, 22(6):42–51.

Palanca, J., Heras, S., Botti, V., and Julian, V., 2014. receteame.com: a Persuasive Social Recommendation System. In *12th International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer.

Phanich, M., Pholkul, P., and Phimoltares, S., 2010. Food recommendation system using clustering analysis for diabetic patients. In *Information Science and Applications (ICISA), 2010 International Conference on*, pages 1–8. IEEE.

Schall, D., 2015. Social Network-Based Recommender Systems.

Teng, C.-Y., Lin, Y.-R., and Adamic, L. A., 2012. Recipe recommendation using ingredient networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 298–307. ACM.

Ueda, M., Takahata, M., and Nakajima, S., 2011. User's food preference extraction for personalized cooking recipe recommendation. *Semantic Personalized Information Management: Retrieval and Recommendation SPIM 2011*, page 98.

Zhou, X., Xu, Y., Li, Y., Josang, A., and Cox, C., 2012. The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37(2):119–132.