Attractiveness Analysis for Health claims on Food Packages

Xiao Li[1], Huizhi Liang[2], Chris Ryder[3], Rodney Jones[3], and Zehao Liu[3]

[1] University of Aberdeen, United Kingdom
x.li.12@aberdeen.ac.uk
[2] Newcastle University, United Kingdom
huizhi.liang@newcastle.ac.uk
[3] University of Reading, United Kingdom
{c.s.ryder,r.h.jones,zehao.liu}@reading.ac.uk

Abstract. Health Claims (Health Claims) on food packages are statements used to describe the relationship between the nutritional content and the health benefits of food products. They are popularly used by food manufacturers to attract consumers and promote their products. How to design and develop NLP tools to better support the food industry to predict the attractiveness of health claims has not yet been investigated. To bridge this gap, we propose a novel NLP task: attractiveness analysis. We collected two datasets: 1) a health claim dataset that contains both EU approved Health Claims and publicly available Health claims from food products sold in supermarkets in EU countries; 2) a consumer preference dataset that contains a large set of health claim pairs with preference labels. Using these data, we propose a novel model focusing on the syntactic and pragmatic features of health claims for consumer preference prediction. The experimental results show the proposed model achieves high prediction accuracy. Beyond the prediction model, as case studies, we proposed and validated three important attractiveness factors: specialised terminology, sentiment, and metaphor. The results suggest that the proposed model can be effectively used for attractiveness analysis. This research contributes to developing an AI-powered decision making support tool for food manufacturers in designing attractive health claims for consumers.

Keywords: Attractive Analysis · Health Claims · Learning-to-Rank · Consumer Preference Prediction.

1 Introduction

Health claims on food packages (e.g. Figure 1) are statements used to describe the relationship between the nutritional content and the health benefits of food products for product promotion. Food manufacturers are increasingly including health claims on their packages [10]. Recent research shows that the presence of such claims on packages generally has a positive impact on consumers' perceptions of the healthiness of products and their willingness to buy them [1]. To protect consumers from being deceived or misled, health claims are strictly regulated in most places in the world. In the European Union, the use of health claims on food packages and in other marketing materials is governed by Regulation (EC) No. 1924/2006. According to the regulation, manufacturers

(a) Approved Health Claims	(b) Revised Health Claims for real commercial use
Vitamin A contributes to the normal function of the immune system	High in vitamin A which supports the normal function of the immune system
Vitamin C contributes to the protection of cells from oxidative stress	Vitamin C contributes to antioxidant activity in the cells (to help protect them from damaging oxidative stress)
Potassium contributes to the maintenance of normal blood pressure	Potassium plays role in maintaining normal blood pressure
Selenium contributes to the normal function of the immune system	Selenium helps maintain your immunity system

Table 1: Example Health Claims approved by Regulation (EC) 432/2012 and example Health Claims collected from food packages

that sell their products within the European Union may only include health claims that have been approved for use by the European Commission based on the verification of their scientific substantiation by the European Food Safety Authority (EFSA). However,

Fig. 1: An example health claim on a food package.

health claims approved by EFSA are written in dense scientific language which is sometimes difficult for consumers to understand. Rewording of approved health claims is allowed as long as the revised claim has the same meaning as the approved claim, as stated in Regulation (EC) 432/2012. In practice, food manufacturers often attempt to rewrite approved health claims in order to communicate the health benefits of their products in an easy-to-understand, unique, and attractive way. Table 1 shows examples of approved and revised health Cclaims. Approved health claims (see Table 1 (a)) consist of information about 1) the nutrient contained in the product, and 2) the health benefits of the nutrient, following a relatively standard template, for example:

[nutrient] *contributes to* **[health benefit]**

Table 1 (b) shows examples of revised health claims. We can see that manufacturers usually revise health claims by by focusing on certain linguistic features, for example, word choice, syntactic structure, emotional valiance, etc. Traditionally, nutritionists, marketers, and lawyers hired by food companies work together to formulate and vet these revised versions of health claims . Then they often conduct user studies or A/B

testing for each revised version. Currently, there is no systematic research attempting to link the attractiveness of health claims to specific linguistic characteristics.

In this paper we propose to analyze the attractiveness of Health Claims on food packages by developing a computational prediction indicator to determine how attractive a health claim is likely to be for consumers. Specifically, we compute a Consumer Preference score (CP score) for specific health claims. The higher the score, the more likely the health claim will be welcomed by consumers.

The challenges of this study include: (1) The difficulty of defining linguistic criteria to measure the attractiveness of health claims. Thus, one cannot evaluate health claims based on pre-defined guidelines. (2) The range of linguistic variables involved. Since EU regulations stipulate that the literal meaning of revised health claims should be similar or the identical to corresponding approved health claims, semantic meaning is unlikely to be the main factor in determining whether or not a revised health claim is more or less attractive than its corresponding approved claim. Other factors including syntax and pragmatics are likely to be more important. (3) The difficulty of obtaining negative labels (i.e., very unattractive health claims), since the revised health claims on food packages have been designed by experts with the explicit aim of attracting consumers. (4) The difficulty of obtaining human evaluations of health claims that substantially diverge from the approved claims, since presenting unapproved or inaccurate health claims to people might be considered unethical.

To address the above challenges, we first collected two datasets, which we discuss in §3.1. Then we designed a consumer preference prediction model, as discussed in §3.2. The evaluation of the proposed model is given in §4. Finally, in §5, we analyse the attractiveness factors of health claims through three case studies based on our model. The contribution of this work can be summarised as follows: (1) We have framed a new application area or task for NLP techniques: attractiveness analysis for health claims in the food industry; (2) We have collected the first health claim datasets for attractiveness analysis; (3) We propose a novel model for studying the attractiveness of health claims, which has achieved high accuracy in our evaluation; (4) We have demonstrated effective means of investigating attractiveness factors in health claims using the model.

2 Related Work

Although health claims are widely used in the food industry, there is surprisingly little scientific research available to support food companies in making decisions about how to formulate Health Claims. Previous research, [17] finds that consumers prefer food products with health claims on the packaging compared to products without health claims, and give greater weight to the information mentioned in health claims than to the information available in the Nutrition Facts panel. [18] suggests that consumers have individual differences in their preferences for health claims, but these differences can also be attributed to consumer cultural differences. Regarding the content of health claims, [9] try to improve health Claims by adopting the Decisions Framing [19]; a study by [5] states that foods that emphasise healthy positive contributions to life (referred to as life marketing) are more attractive to consumers than foods that emphasise avoidance of disease (death marketing); [20] finds that consumers may be reluctant to try products

whose health claims include unfamiliar concepts because consumers tend to evaluate them as less credible. However, all of this research focuses on the relationship among food products, health claims, and consumer attitudes rather than focusing on the the specific linguistic features of health claims.

Currently, NLP techniques have been widely applied in linguistic studies [8]. NLP allows us to use quantitative research methods to study abstract linguistic phenomena. NLP is particularly popular in studeis of syntax and pragmatics [3], e.g., dependency parsing [13], metaphor processing [15], and sentiment analysis [6]. However, to the best of our knowledge, applying these NLP tools to the analysis of the attractiveness of health claims is still new.

3 Consumer Preference Prediction of Health Claims

Learning consumer preferences for health claims is defined as learning a function ($f(\cdot)$) that maps health claims to a real number (called the Consumer Preference score or CP score for short, denoted by u). The real number indicates the degree of consumer preference for an input health claim text (denoted by x). The larger the value of u, the more attractive the health claim is assumed to be.

$$f(\mathrm{x}) = u \tag{1}$$

3.1 Dataset Collection

Our dataset collection had two stages, which generates a health claims dataset with pair-wise customers preference labels. In the first stage, we collected a large number of real-life health claims from the food products sold in EU supermarkets. Since vitamins and minerals are everyday nutrients that are familiar to consumers and can be found in many food products, we only used the Health Claims for vitamins and minerals in our research to better control the experimental variables. At this stage, we collected a total of 4200 Health Claims text.

In the second stage, we used scenario-based experiments to observe consumers' (virtual) purchase intentions, based on the collected Health Claims. Figure 2 shows an example task in our experiments. In the experiments, a subject is asked to help Alex to choose a food product as a gift for Sam. We used neutral names (Alex and Sam) and pronouns "them" rather than "he" or "she" to avoid gender bias. The decision was made according to two random-paired health claims. Each subject was asked to complete 20 tasks. The options in each task are displayed as a gift pair consisting of the random health claims that we collected. Because of the randomness, the questionnaires of one subject might be different from one another. We recruited 200 subjects from the EU and the UK[4] via Amazon MTurk. 183 of them were valid, i.e. those who completed the questionnaire within 30 minutes and spent more than 3 seconds on each task on average. Totally, we gathered 183x20=3660 answers for the paired health claims with a selected preferred label for each pair.

[4] The locations of subjects were filtered by MTurk.

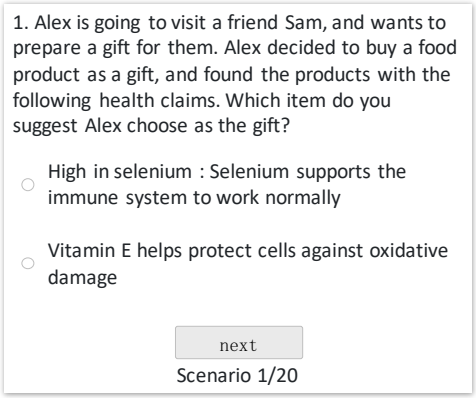


Fig. 2: A scenario task example for the Consumer Health Claim Preference data collection.

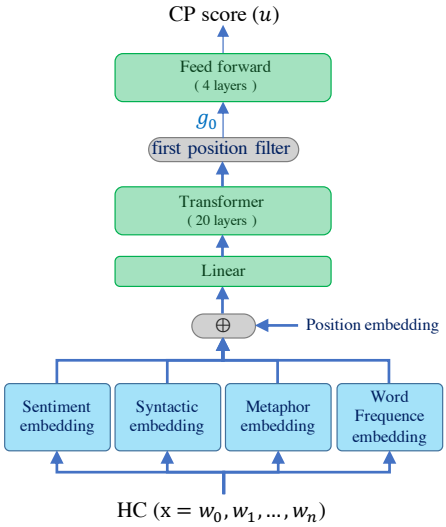


Fig. 3: The architecture of our proposed model (i.e. $f(\cdot)$ in Eq. 1).

3.2 Prediction Model

Our model is used to learn $f(\cdot)$ in Eq. 1. The input of the model is a health claim, and the output is the Consumer Preference score of the health claim. The model aims to focus as much as possible on the preferences for linguistic features in health claims rather than their literal meanings. First, as the literal meanings of revised health claims are governed by EU regulations, revised health claims should mainly have the same semantic meaning as the corresponding approved health claims. Second, previous research [5] suggests that consumer preferences for health claim are affected by deep-level linguistic feature,

such as sentiment factors [5] or unfamiliar concepts [20]. If a model focuses on specific words but ignores the overall Health Claim sentence, it may lack practical significance.

The model adopts a Transformer [21] based architecture (see Figure 3), since it has demonstrated its advantage in many NLP tasks. The input is a health claim sentence with a special token ([sos]) placed at the beginning of the health claim sentence. Unlike common NLP practices that mainly embed semantic information of the input text in vector space, we considered multiple factors focusing on syntax and pragmatic factors.

Specifically, we encoded tokens (i.e. words, punctuation etc.) denoted as $(w_0, ..., w_n)$ of the input as vectors $(h_0, ..., h_n)$ via a specially designed embedding layer which consists of multiple pre-trained models in syntactic and pragmatic analysis tasks [7]. The outputs of these pre-trained models were concatenated together as the input of the transformers of our model, as shown in Figure 3. First, we used the model of [7] to calculate the sentiment scores of each word. The outputs of this sentiment model are 4-dimensional vectors including the overall sentiment score (in $[-1, 1]$), the positive, neutral, and negative scores (in $[0, 1]$). The model of [14] was employed to generate metaphoricity scores that indicate the metaphoric possibility for each word (i.e., a real number in $[0, 1]$).

Second, we parsed health claims with a python NLP module (spaCy) to obtain the dependency relationship features for forming the syntactic embedding in Figure 3. For each word, we adopted the distance between a word and its dependent word as the partial syntactic information. Here, we encode the distances by using the same method of the learnt positional embedding. For example, suppose w_1 depends on w_2 and w_2 depends on w_6 which means the distances are $2 - 1 = 1$ and $6 - 2 = 4$, we encode 1 and 4 by the way of the positional embedding as the partial embedding for w_1 and w_2. The edge tags (i.e. grammatical relations annotating the dependencies e.g. *dobj* for direct objects, *conj* for conjunct, etc.) were also employed to represent the syntactic relationship, which was encoded with one-hot encoding.

Finally, we used the logarithm of word frequency and the tf-idf values as word frequency embedding features to assign more weight to rare words and concepts. The word frequencies and the document frequency were obtained from Google Books Ngram dataset [5]. We also used the learnt position embedding to identify the word positions. The final embedding vector (h_i for word w_i) is the concatenation of all the 5 kinds of embedding features. The embedding vector for [sos] token was set to a zero vector. After the embedding layer, $h_0, ..., h_n$ were fed into a linear layer to compress their feature maps to 32-dimension, before feeding into a 20-layer transformer encoder. According to [2], narrowing the width (32-dimension) of a model and increasing its depth (20-layer) can mitigate overfittings. Then, the output vector for the special token [sos] was used as the pooling hidden vector (g_0), feeding to a 4-layer feed forward neural network, obtaining the final Consumer Preference score (u).

The training process adopted a Learning-to-Rank strategy [4] for learning global sorting scores (i.e. the Consumer Preference score u) of health claims from paired examples. Given two paired health claims (x_r^+, x_r^-), where x_r^+ is preferred to x_r^- in an health claim pair (r). The training process is to force the Consumer Preference score of x_r^+ to be higher than that of x_r^-, that is to let $u_r^+ > u_r^-$ (Eq. 2).

[5] https://storage.googleapis.com/books/ngrams/books/datasetsv3.html

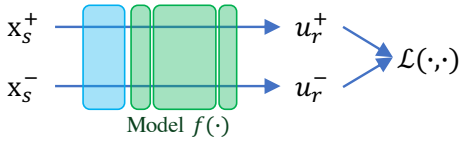


Fig. 4: The training process (Eq. 3). Two paired health claims are fed to the model iteratively, where the model learns to yield a higher score (u_r^+) for the preferred health claim than the score of the other health claim (u_r^-) of a given health claim pair.

$$\begin{aligned} u_r^+ &= f(\mathrm{x}_r^+) \\ u_r^- &= f(\mathrm{x}_r^-) \end{aligned} \tag{2}$$

Unlike the usual practices of Learning-to-Rank, our model adopted an exponential loss function (Eq. 3) R denotes the training dataset and $|R|$ is the size of R. It forces the value of $u_r^- - u_r^+$ to be as small as possible just as the usual Learning-to-Rank approaches. However, when $u_r^- - u_r^+ > 0$ (i.e., wrong predicted preference order of x_s^+ and x_s^-), our exponential loss yields extra penalties.

$$\mathcal{L} = \frac{1}{|R|} \sum_{r \in R} e^{u_r^- - u_r^+} \tag{3}$$

In the training process, x_s^+ and x_s^- in a paired health claims are fed into the model iteratively, so that when the model predicts Consumer Preference score for one health claim, another health claim is not considered (Figure 4). According to the loss (Eq. 3), the gradients on the model depend on the relative Consumer Preference scores (i.e. $u_r^- - u_r^+$) rather than the individual values of u_r^- and u_r^+. If and only if the model predicts the wrong order for two health claims (x_r^+ and x_r^-), the model parameters will be updated significantly.

4 Evaluation and Results

We examined our model with 50-cross validation. Random 80% of data was used for training, and 20% was used for testing. The model was trained for 500 epochs with a batch size of 256. The accuracy of automatic evaluation was measured by the averaged accuracy of the last 10 epochs. We used an AdamW optimiser with the learning rate of 10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$, where the weight decay was 10^{-2}. The example health claims with the predicted Consumer Preference scores are ranked in Table 3. We introduced a baseline model with a learnt word embedding layer with randomly initialised weights layer (named Learnt embedding in Table 2), and a BERT baseline (named Bert). We allowed weight updating for Bert and the Learnt embedding layer during training (Bert with fine tune). Since we proposed the exponential-based loss function, we also compared it with the original Learn-to-Rank loss which is the Cross-Entropy Loss (i.e. the model adopting our embedding method and Cross-Entropy Loss). We tried to test the utilities of our embedding layer.

As seen in Table 2, our proposed model achieved an accuracy of 76% on the testing set, outperforming the baselines by at least 8%. This can be explained by the fact that the pre-trained models in our embedding layer provided extra syntactic and pragmatic knowledge. Reflecting on the accuracy scores, the BERT baseline yielded a high score on the training set but low on the testing set. We infer that the BERT model exposes overfitting on the training set. This can be explained by the fact that BERT provides rich semantic features, but they may not be suitable for the attractiveness analyses of Health Claims. In addition, the Learnt embedding baseline seems not to have learnt enough appropriate knowledge for the prediction. This is probably due to the fact that the size of our training data set was not large enough for this kind of model.

	Training set accuracy	Testing set accuracy
Learnt embedding	0.86	0.66
Bert	0.95	0.68
CrossEntropy loss	0.83	0.73
Proposed model	0.84	**0.76**

Table 2: The automatic evaluation results for the consumer preference prediction. The accuracy is the mean of the accuracy scores of 50-cross-validation.

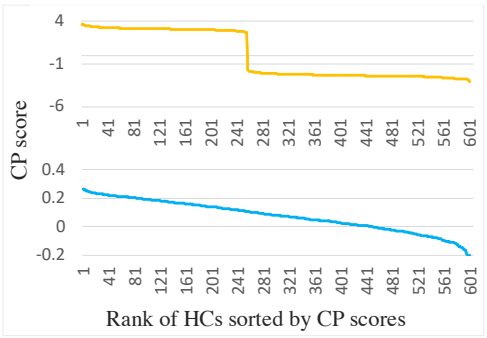


Fig. 5: The changing of the Consumer Preference scores on the sorted Health Claims. Yellow line denotes that of Cross Entropy Loss; Blue line denotes that of our exponential loss.

We also compared the influence of our proposed loss function (Eq 3) with the original Learn-to-Rank loss (i.e. the Cross-Entropy Loss). By respectively using the proposed model and the baseline model using Cross-Entropy loss (in Table 2), we scored all the Health Claims and sorted them according to the scores. After that, we investigated how the scores changed (Figure 5) when the health claim rank changed. We can see that when

Rank	CP score	Health Claims
1	0.266	It is a source of phosphorus: Phosphorus helps ensure the normal energy metabolism
2	0.261	Vitamin C helps to support a healthy immune system
3	0.260	High in vitamin B12: Vitamin B12 supports the immune system to function normally
4	0.259	High in vitamin A: Vitamin A supports the immune system to function normally
5	0.257	Naturally high in vitamin C: Vitamin C helps to protect cells from oxidative stress
...	...	...
597	-0.199	Calcium contributes to normal muscle function
598	-0.199	Magnesium contributes to normal muscle function
599	-0.199	Zinc contributes to normal macronutrient metabolism
600	-0.200	Potassium contributes to normal muscle function
601	-0.202	Zinc contributes to normal cognitive function

Table 3: The top 5 Health Claims and the bottom 5 Health Claims, sorted by predicted Consumer Preference scores. Empirically, we could see that the top 5 Health Claims are much more attractive than the bottom 5 Health Claims.

Hypothesis	Group	Ratio of chosen	p-value
$G_H \succ G_L$	G_H	.82	$< 10^{-5}$
	G_L	.18	

Table 4: The human evaluation results against null hypothesis $\mathbf{H_G}$. The p-values are calculated via Mann-Whitney U test.

the model is trained with our exponential loss, the Consumer Preference scores show a smooth change. While it is trained by the Cross-Entropy loss, there is a sharp drop in the Consumer Preference scores. This drop may cause the model with Cross-Entropy loss to perform slightly worse than the proposed model.

In addition to the automatic evaluations, we also conducted human evaluations. We collected the top 50 health claims (denoted by G_H) that had the highest Consumer Preference scores and the last 50 health claims with the lowest scores (denoted by G_L). We conducted human evaluation by pairing each health claim in G_H with a randomly selected health claim in G_L. We meant to test whether the health claims with high Consumer Preference scores (G_H) were more preferred than the health claims with low Consumer Preference scores (G_L) via human evaluations. Based on the same experimental method in §3.1, 20 subjects participated in this survey, and each of them completed 8 tasks. Our null hypothesis ($\mathbf{H_G}$) was that health claims in G_H and G_L have no difference regarding consumer preferences. The results (Table 4) show the subjects significantly prefered (denoted by '$\succ$') the top-ranking health claims (G_H) compared to the low-ranking health claims (G_L); the p-value $< 10^{-5}$ rejects $\mathbf{H_G}$. Thus, a health claim with a high Consumer Preference score is more likely preferred by a consumer than a low Consumer Preference score health claim.

5 Case Studies

This section describes our investigation into the linguistic factors affecting the attractiveness of a health claim through case studies. We conducted two kinds of case study from the perspectives of local factors and global factors, which was to: (1) understand the important factors that determine whether a health claim is attractive or not; (2) verify whether the results, given by our model, are consistent with the consumer survey results. The findings will help manufacturers to automatically evaluate their health claims before marketing.

5.1 Specialised Terminology Factors

[20] suggests that unfamiliar concepts may prevent consumers from buying food products. [16] found that the use of specialised terminology can reduce the number of citations of papers in the domain of cave research. In light of this, we analysed the impact of specialised terminology (jargon) in health claims. Specialised utterances are local features for sentences; they are mainly related to the academic names of nutritional ingredients. E.g., *thiamin* is also known as *Vitamin B1*. The following example shows two health claims with different nutrient terminologies.

***Thiamin** helps support a healthy heart*

Vitamin B1 helps support a healthy heart

We focused on B vitamins, e.g., *thiamin* vs. *Vitamin B1*, *riboflavin* vs. *Vitamin B2*, and *pantothenic acid* vs. *Vitamin B5*, because they can be easily found in food products. Table 5 shows the statistics of each item in our collected health claim dataset.

Specialised utterance		Common utterance	
Utterance	count	Utterance	count
thiamin	25	vitamin B1	32
riboflavin	7	vitamin B2	14
pantothenic acid	14	vitamin B5	14

Table 5: The statistics of the specialised names and common names for B vitamins in our collected dataset.

First, we tested the Consumer Preference score differences between using specialised and common utterances. We extracted all 46 health claims containing the specialised utterances (see Table 5), denoted as V_s. We developed another set, $V_{s\to c}$, by replacing the specialised utterances of health claims in V_s with the corresponding common utterances. Similarly, a set of 60 health claims containing the common utterances is denoted as V_c, while $V_{c\to s}$ consists of the health claims in which the common utterances in V_c are replaced with the corresponding specialised utterances. We computed the Consumer Preference scores for each health claim in V_s, $V_{s\to c}$, V_c and $V_{c\to s}$ respectively.

Next, we compared the Consumer Preference score differences between vitamins and minerals since minerals have no alternate name used in the food industry. Although vitamins and minerals belong to different nutrient categories, since vitamins are more common than mineral names in daily life, the latter is more obscure than the former for consumers. In line with the above method, we collected all 354 health claims containing vitamins, such as 'Vitamin A', 'B1' and 'C', developing a health claim set G_v. Its paired set $G_{v\to m}$ is developed by replacing a vitamin with a random mineral. Similarly, we gathered a mineral set G_m that has 191 health claims, and its corresponding vitamin set $G_{m\to v}$. We computed the Consumer Preference scores for G_v, $G_{v\to m}$, G_m and $G_{m\to v}$ respectively. The results are shown in Table 6a. By replacing specialised items with their corresponding common names, a large proportion of health claims (76%) achieved higher Consumer Preference scores. On the other hand, if common items were replaced with specialised names, most of the Consumer Preference scores (72%) decreased. This demonstrates that consumers prefer the common names of vitamins to their academic names in health claims. The comparison between vitamins and mineral shows that changing minerals to vitamins can yield higher Consumer Preference scores, while the reverse brings negative impacts. Thus, consumers prefer vitamins to minerals in health claims. This can be explained by the fact that consumers prefer common or familiar concepts in health claims on food in their daily life.

We further verify these statistical findings with human evaluation. We randomly select Health Claims from V_s, V_c, G_v, and G_m, respectively. Health Claims from V_s and V_c are paired for evaluating, when the survey is to test whether the common utterances are preferred to the specialised utterances. Similarly, health claims from G_v and G_m are paired to test whether vitamins are preferred to minerals. The survey was conducted based on the same method described in §3.1. We gathered 120 valid answers (6 tasks per person) from 20 subjects via Amazon MTurk. The statistical results are shown in Table 6b. Just as with the statistical findings, the human evaluation supports the argument that specialised utterances may reduce the attractiveness of health claims. Compared with conducting human evaluation, our automatic attractiveness analysis model is more efficient and simpler.

	CP score ↑		C P score ↓	
	count	ratio	count	ratio
$V_s \to V_{s\to c}$	35	**.76**	11	.24
$V_c \to V_{c\to s}$	17	.28	37	**.72**
$G_v \to G_{v\to m}$	103	.29	251	**.71**
$G_m \to G_{m\to v}$	163	**.85**	28	.15

(a) Changes of Consumer Preference scores by using alternative names (V) and different nutrients (G). The p-values are less than 0.0005 based on Mann-Whitney U test.

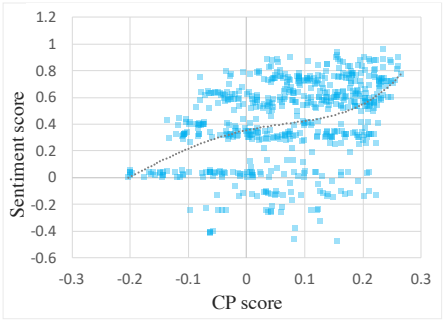
Hypothesis	Group	Ratio of chosen	p-value
$V_s \succ V_c$	V_s	.63	$< .006$
	V_c	.37	
$G_v \succ G_m$	G_v	.60	$< .029$
	G_m	.40	

(b) The human evaluation results to compare alternative names and different nutrients. The p-values are calculated via Mann-Whitney U test.

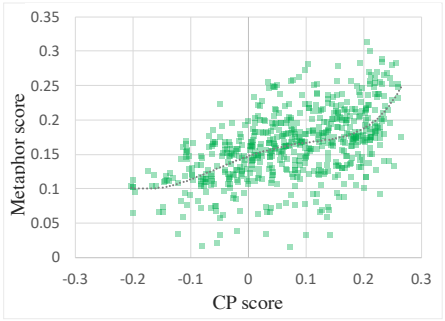
Table 6: Changes of Consumer Preference scores and human evaluation results

5.2 Sentiment and Metaphoricity Factors

This section investigates the influence of sentiment and metaphoricity features, which are the global features for sentences. For example, *"Natural source of vitamin A : contributing to boost your immune system"* has positive sentiment and contains a metaphor. Both sentiment and metaphoricity features are pragmatic features. health claims with different levels of sentiment polarities could emotionally impact consumer decisions. We computed the Pearson Correlation Coefficient between Consumer Preference scores and sentiment scores for all the health claims that we collected in §3.1. Sentiment scores are given by the model of [7] (here we only use the overall sentiment score). We visualise the distribution of the sentiment scores by Consumer Preference score in Figure 6a. The moderate correlation coefficient (0.40 with p-value $< 10^{-5}$) suggests that health claims with positive sentiment are more attractive than negative Health Claims.



(a) Visualisation of the correlation for Consumer Preference score (CP score) and sentiment score among the health claim collection with an order 4 polynomial trendline. Their Pearson Correlation Coefficient is 0.40 with p-value $< 10^{-5}$.



(b) Visualisation of the correlation for Consumer Preference score (CP score) and metaphoricity score among the health claim collection with an order 4 polynomial trendline. Pearson correlation coefficient is 0.52 with the p-value $< 10^{-5}$.

Fig. 6: Sentiment and Metaphor

Metaphoricity is another pragmatic feature. Metaphorical expressions convey information conceptually [11], because they use one or several words to represent a different concept, instead of the original literal concepts. Thus, consumers may receive richer information from metaphoric health claims. Similar to sentiment features, we studied the correlation between Consumer Preference scores and metaphoricity scores. The metaphoricity scores are given by the model of [14], whose values are in a range of $[0, 1]$. The higher the score, the more likely the health claim is metaphoric. The correlation coefficient score (0.52 with p-value $< 10^{-5}$) suggests that consumers likely prefer health claims that use metaphoric expressions.

The above statistical analysis signifies that using words with positive sentiment polarities and metaphoric language in health claims can attract more consumers. We

Hypothesis	Group	Ratio of chosen	p-value
$S_H \succ S_L$	S_H	.58	$< .003$
	S_L	.42	
$M_H \succ M_L$	M_H	.55	$< .042$
	M_L	.45	

Table 7: The human evaluation results of sentiment and metaphor factors. The p-values are calculated via Mann-Whitney U test.

conducted a human evaluation to verify the hypothesis that health claims with higher sentiment/metaphor scores are more attractive. We paired the health claims with high sentiment scores (S_H with a sentiment score above 0.5) and health claims with low sentiment score (S_L with a sentiment score below -0.5). We also paired the high metaphoricity score health claims (S_H with a metaphoricity score above 0.2) with low metaphoricity score health claims (S_L with a metaphoricity score below 0.1). The human evaluation followed the same process as in §4.1. We recruited 20 valid subjects and gathered 400 answers from them. As seen in Table 7, the human evaluation results support our hypothesis. In practice, one may choose positive lexicons or use metaphoric expressions in writing health claims, which attract consumers.

6 The deployment of the proposed attractiveness analysis model

The proposed attractiveness analysis model is one major component of our funded project.[6] [7] [8]. The project had two components: a consumer toolkit and a manufacturer platform (called "Research, Analytics, and Consumer Insights Platform"). The consumer toolkit provides multiple interactive online *activities* including educational activities, and practice activities to teach health claim knowledge to consumers. This toolkit also tests and collects users' data about their understanding of the attractiveness of health claims. The manufacturer platform aims to support food manufacturers to evaluate their created health claims. The proposed attractiveness analysis model is used as the prediction engine of the manufacturer platform. By learning the consumer preferences from the collected data, there are two NLP-based prediction models in the manufacturer platform. The proposed attractiveness analysis model predicts how much consumers might like the Health Claims with 5-scale scores for two different scenarios. The first model predicts the general consumer preference score, which reflects the preference of the population. The second model is a conditional model, which predicts the target consumer preference scores – the preference of consumer characteristics (e.g. gender, age, etc.) is specified by the platform users. A detailed description of the design and implementation of the Consumer Toolkit is shown in our previous published system demonstration paper [12].

For the given health claims, the manufacturer platform can show the general consumer preference scores for the population, and the target consumer preference scores for

[6] Project website: https://www.healthclaimsunpacked.co.uk/
[7] Consumer Toolkit website: https://www.unpackinghealthclaims.eu/
[8] Manufacturer Platform website: https://www.healthclaimsinsights.eu/

Fig. 7: The home page of the consumer toolkit

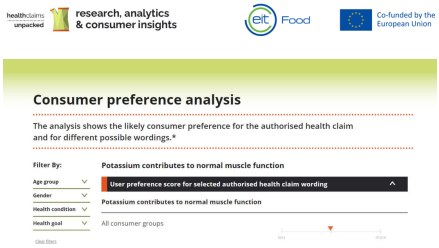


Fig. 8: The home page of the manufacture platform

groups of consumers by their characteristics (e.g. age-based groups, gender-based groups etc). It suggests different wordings for the query health claim with the same nutrient and health benefit. The predicted consumer preference scores of the suggested health claims are shown to the users, which helps users to make decisions. A prototype of the consumer toolkit was released in English in November 2019. Versions of the consumer toolkit in five other European languages including German, French, Polish, Romanian and Hungarian were released in December 2021. Data from the consumer toolkit were used as a foundation to develop the manufacturer platform. The manufacturer platform was released in early 2022. It only has English version so far. It has the potential to be extended for predicting the attractiveness of health claims in other languages.

7 Conclusion

This paper discussed how to apply NLP techniques to better support the food industry for attractiveness analysis of health claimss. By introducing the new NLP task – Attractiveness Analysis, we developed a novel model to predict the consumer preferences of health claims. The model was trained on a newly collected health claim dataset with an improved Learn-to-rank loss function. By explicitly focusing on the syntactic and pragmatic features, the model successfully predicts consumer preference with high accuracy. Based on this model, we investigated and validated three important attractiveness factors. We observed that using common names instead of specialised academic names in health claims is more attractive. In addition, positive and metaphoric lexicons are also preferred. Our model can help manufactures evaluate their health claims without conducting a human-based survey. We also discussed the deployment of the proposed model in the manufacturer platform of the project system. In the future, we will explore a data-driven approach to identify more attractiveness factors and develop automatic attractiveness analysis tools for multi-lingual health claims. Also, we will consider the legal requirements and explore novel NLP tools to support food manufactures in designing health claims that are both attractive and legitimate (i.e., not deviating from the meaning of the original EFSA approved claim).

8 Acknowledgements

This project is funded by European Institute of Innovation and Technology (EIT) Food/EU Horizon 2020 (project number 19098). The authors would like to thank the funder EIT food and the great help and support of other team members of the project.

References

1. A. Saba, e.a.: Country-wise differences in perception of health-related messages in cereal-based food products. Food Quality and Preference **21**(4), 385–393 (2010)
2. Bello Irwan, e.a.: Revisiting resnets: Improved training and scaling strategies. arXiv preprint arXiv:2103.07579 (2021)
3. C. Erik, e.a.: Jumping nlp curves: A review of natural language processing research. IEEE Computational intelligence magazine **9**(2), 48–57 (2014)
4. Cao Zhe, e.a.: Learning to rank: from pairwise approach to listwise approach. In: ICML'07. pp. 129–136 (2007)
5. Euromonitor: Functional foods — a world survey (2020)
6. Hussein, D.M.E.D.M.: A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences **30**(4), 330–338 (2018)
7. Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 8 (2014)
8. Jurafsky, D.: 26 pragmatics and computational linguistics. The handbook of pragmatics p. 578 (2004)
9. Krishnamurthy P., e.a.: Attribute framing and goal framing effects in health decisions. Organizational behavior and human decision processes **85**(2), 382–399 (2001)
10. Lähteenmäki, L.: Claiming health in food products. Food Quality and Preference **27**(2), 196–201 (2013), ninth Pangborn Sensory Science Symposium
11. Lakoff, G., Johnson, M.: Metaphors We Live by. University of Chicago press (1980)
12. Li, X., Liang, H., Liu, Z.: Health Claims Unpacked: A Toolkit to Enhance the Communication of Health Claims for Food, p. 4744–4748. Association for Computing Machinery, New York, NY, USA (2021), https://doi.org/10.1145/3459637.3481984
13. Li, Z., Cai, J., He, S., Zhao, H.: Seq2seq dependency parsing. In: ICCL'18. pp. 3203–3214 (2018)
14. Mao, R., Li, X.: Bridging towers of multitask learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (2021)
15. Mao, R., Lin, C., Guerin, F.: Word embedding and WordNet based metaphor identification and interpretation. In: ACL'18. vol. 1, pp. 1222–1231 (2018)
16. Martínez, A., Mammola, S.: Specialized terminology reduces the number of citations of scientific papers. Proceedings of the Royal Society B **288**(1948), 20202581 (2021)
17. Roe Brian, e.a.: The impact of health claims on consumer search and product evaluation outcomes: results from fda experimental data. Journal of Public Policy & Marketing **18**(1), 89–105 (1999)
18. Tino Bech-Larsen, K.G.G.: The perceived healthiness of functional foods: A conjoint study of danish, finnish and american consumers' perception of functional foods. Appetite **40**(1), 9–14 (2003)
19. Tversky, A., Kahneman, D.: The framing of decisions and the psychology of choice. science **211**(4481), 453–458 (1981)

20. Van Kleef, E., van Trijp, H.C., Luning, P.: Functional foods: health claim-food product compatibility and the impact of health claim framing on consumer evaluation. Appetite **44**(3), 299–308 (2005)
21. Vaswani Ashish, Shazeer Noam, P.N.U.J.J.L.G.A.N.K.L.P.I.: Attention is all you need. In: NIPS (2017)