Separating Brands from Types: an Investigation of Different Features for the Food Domain

Michael Wiegand and **Dietrich Klakow**
Spoken Language Systems
Saarland University
D-66123 Saarbrücken, Germany
{Michael.Wiegand|Dietrich.Klakow}@lsv.uni-saarland.de

Abstract

We examine the task of separating types from brands in the food domain. Framing the problem as a ranking task, we convert simple textual features extracted from a domain-specific corpus into a ranker without the need of labeled training data. Such method should rank brands (e.g. *sprite*) higher than types (*e.g. lemonade*). Apart from that, we also exploit knowledge induced by semi-supervised graph-based clustering for two different purposes. On the one hand, we produce an auxiliary categorization of food items according to the Food Guide Pyramid, and assume that a food item is a type when it belongs to a category unlikely to contain brands. On the other hand, we directly model the task of brand detection using seeds provided by the output of the textual ranking features. We also harness Wikipedia articles as an additional knowledge source.

1 Introduction

Brands play a significant role in social life. They are the subject matter of many discussions in social media. Their automatic detection for information extraction tasks is a pressing problem since, despite their unique property to refer to commercial products of specific companies, in everyday language they often occur in similar contexts as common nouns. A typical domain where such behaviour can be observed is the food domain, where **food brands** (e.g. *nutella* or *sprite*) are often used synonymously with the **food type**[1] of which the brand is a prototypical instance (e.g. *chocolate spread* or *lemonade*). Such usage is illustrated in (1) and (2).

(1) In the evening, I eat a slice of bread with either **nutella** or marmalade.

(2) I prepare my pancakes with baking soda, water and a lacing of **sprite** instead of sugar.

This particular phenomenon of metonymy (Lakoff and Johnson, 1980), commonly referred to as *generici-zed trademarks*, of course, has consequences on automatic lexicon induction methods. If one automatically extracts food types, one also obtains food brands.

In this paper, we examine features to detect brands automatically. Solving the issue with the help of a manually-compiled list of brands neglects parts of the nature of brands. Brands come and go. Some products may be discontinued after a certain amount of time (e.g. due to limited popularity) while, on the other hand, new products constantly enter the market. For instance, popular food brands, such as *sierra mist* or *kazoozles*, did not exist a decade ago. Therefore, a list of brands that is manually created today may not reflect the predominant food brands that will be available in a decade.

The features we introduce to detect brands consider both the intrinsic properties of brands and their contextual environment. Even though in many contexts, brands are used as ordinary type expressions (1), there might be specific contexts that are only observed with brands. We also consider distributional properties: brands may co-occur with other brands. Moreover, they may be biased towards certain categories, e.g. sweets, beverages etc. For the latter, we actually exploit the usage of food brands to be

[1] We define *food type* as common nouns that denote a particular type of food, e.g. *apple*, *chocolate*, *cheese* etc.

Method	Corpus	Corpus Type	P@10	P@100	P@500
ranking by frequency	chefkoch.de	domain specific	0.00	22.00	25.60
induction based on coordination	Wikipedia	open domain	90.00	60.00	47.80
induction based on coordination	chefkoch.de	domain specific	100.00	98.00	92.00

Table 1: Precision at rank n (P@n) of different food induction methods.

Label	Items	Examples
Food Types	1745	apple, baguette, beer, corn flakes, crisps, basmati rice, broccoli, chocolate spread, gouda, orange juice, pork, potato, steak, sugar
Food Brands	221	activia, babybel, becel, butterfinger, kit kat, nutella, pepsi, philadelphia, smacks, smarties, sprite, ramazzotti, tuborg, volvic

Table 2: Gold standard of the food vocabulary.

used as genericized trademarks, allowing food categorization methods for types to be easily extended to brands. Moreover, we examine how external knowledge resources, such as Wikipedia, can be harnessed as a means to separate brands from types. Our task is lexicon construction rather than contextual entity classification, that is, we are interested in what a food item generally conveys and not what it conveys in a specific context.

We consider the food domain as a target domain since there are large, unlabeled domain-specific corpora available gathered from social media which are vital for the methods we explore. It is also a domain for which there has already been done research in the area of natural language processing (NLP), and there are common applications, such as virtual customer advice or product recommendation, that may exploit such NLP technology.

The methods we consider require no, or hardly any human supervision. Thus, we imagine that they can also be applied to other domains at a low cost. In particular, other life-style domains, such as fashion, cosmetics or electronics show parallels, since comparable textual web data from which to extract domain-specific knowledge are available.

Our experiments are carried out on German data, but our findings should carry over to other languages since the issues we address are (mostly) language universal. All examples are given as English translations. We use the term **food item** to refer to the union of food brands and food types. All food items will be written in lowercase reflecting the identical case spelling in German, i.e. types and brands are *both* written uppercase. In English, both types and brands can be written uppercase or lowercase[2], however, there is a tendency in user-generated content/social media to write mostly lowercase.

2 Motivation & Data

Previous research on lexicon induction proposed a widely applicable method based on *coordination* (Hatzivassiloglou and McKeown, 1997; Riloff and Shepherd, 1997; Roark and Charniak, 1998): First, a set of seed expressions that are typical of the categories one wants to induce are defined. Then, additional instances of those categories are obtained by extracting *conjuncts* of the seed expressions (i.e. all expressions that match *<seed> and/or <expression>* are extracted as new instances). A detailed study of such lexicon induction has recently been published by Ziering et al. (2013), who also point out the great semantic coherence of conjuncts.

This method can also be applied to the food domain. As a domain-specific dataset for all our experiments, we use a crawl of *chefkoch.de*[3] (Wiegand et al., 2012) consisting of $418,558$ webpages of forum entries. *chefkoch.de* is the largest German web portal for food-related issues. Table 1 shows the effectiveness of coordination as a means of extracting food items from our domain-specific corpus. Given a seed set of 10 frequent food items (we use: *water*, *salt*, *sugar*, *salad*, *bread*, *meat*, *cake*, *flour*, *butter* and

[2]There are plenty of food types that are written uppercase, e.g. *Jaffa Cakes*, *Beef Wellington*, *BLT*, *Hoppin' John* etc.

[3]`www.chefkoch.de`

Properties	Type of Property	Example	Brands	Types
nonwords	general	ebly, sprite, twix	41.63	-*NA*-
derived from proper noun	general	cheddar, evian, jim beam	31.22	2.29
foreign words	general	camembert, merci, wasabi	27.15	12.37
length	general	*average no. of characters*	7.97	10.53
word initial plosives	stylistic	p,t,k,b,d,g (*attract attention*)	31.22	35.81
assonance	stylistic	f**a**nt**a**, k**i**w**i** (fruit), p**a**p**a**y**a**	11.76	11.06
alliteration	stylistic	**b**a**by**b**e**l, **b**lue**b**erry, **t**ic **t**ac	6.79	3.78
onomatopoeia	stylistic	**crunch**ips, **pop**corn	2.71	0.52
rhyme	stylistic	j**elly** b**elly**, h**ubba** b**ubba**	1.35	0.06

Table 3: Comparison of intrinsic properties between brands and types; brands are always underlined; all numbers (except for *length*) are the proportion with the respective property.

potato), we compute all conjuncts and rank them according to frequency. We do this on our domain-specific corpus and on *Wikipedia*. As a baseline, we simply sort all nouns according to frequency in our domain-specific corpus. The table shows that ranking by frequency is no effective method. Conjuncts produce good results provided that they are extracted from a domain-specific corpus.

Even though coordination is a very reliable method to induce food items, it fails to distinguish between food types and food brands. We produced a labeled food vocabulary to be used for all our subsequent experiments consisting of food types and food brands (see Table 2). The food types exclusively comprise the food vocabulary from Wiegand et al. (2014). The food brands were manually selected with the help of the web. We only include food items that occur at least 5 times in our corpus. In our food vocabulary, 87% of our food brands occur as a conjunct of a food type. Therefore, the problem of confusing brands with types is inherent to induction based on coordination.

3 Intrinsic Properties

Table 3 provides some statistics on intrinsic properties of our food items giving some indication which feature types might be used for this task. We also include some stylistic properties of brands that have been addressed in previous marketing research and applied psychology. We focus on fairly straightforward features from *desirable brand name characteristics* (Robertson, 1989), since we assume that there is more general agreement on the underlying concepts than there is on the concepts underlying complex sound symbolism (Klink, 2000; Yorkston and Menon, 2004). For the statistics in Table 3, most properties (i.e. all except *length* and *word-initial plosives*) have been detected manually. The reason for this is that their automatic detection is not trivial (e.g. there is no established algorithm to detect onomatopoeia; even the detection of rhyme or assonance is not straightforward given the low grapheme-phoneme correspondence of English). We did not want the statistics for this exploratory experiment to be distorted by error-proneness of the detection methods.

Table 3 shows that a large part of brands are nonwords indicating that this task is hard to be solved with intrinsic features only. Since there is a high number of brands that are derived from some *existing* proper noun being either a person or a location, named-entity recognition might be applied to this task. Many brands are also foreign words. Unfortunately, applying language checking software on our food items turned out to perform poorly. (These tools are only effective on longer texts, e.g. sentences or entire documents, and do not work on isolated words, as in our problem setting.) We also noticed a difference in average word length between brands and types which is consistent with Robertson (1989) who claims that brand names should be *simple*. Most stylistic features seem to be less relevant to our task as they are either too infrequent or not discriminative. Therefore, we do *not* consider them as features for the detection of brands in our forthcoming experiments.

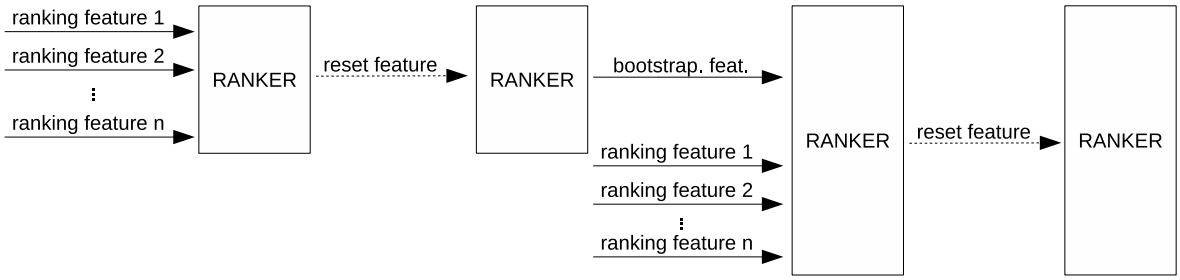

Figure 1: Processing Pipeline for ranking.

Feature Type	**Features**
ranking feature	LENGTH, COMMERCE, NER$_{target}$, NER$_{context}$, DIVERS, PAT$_{mod}$, PAT$_{pp}$
reset feature	GRAPH$_{pyramid}$
bootstrapping feature	GRAPH$_{brand}$, WIKI, VSM

Table 4: Feature classification.

4 Method

Our aim is to determine predictive features for the detection of brands. Rather than employing some supervised learner that requires manually labeled training data, we want to convert these features directly into a classifier without costly labeled data. We conceive this task as a ranking task. The reason for using a ranking is that our features can be translated into a ranking score in a very straightforward manner. For the evaluation, we do not have to determine some empirical threshold separating the category *brand* from the category *type*. Instead, the evaluation measures we employ for ranking implicitly assume highly ranked instances as brands and instances ranked at the bottom as types.

For the ranking task, we employ the processing pipeline as illustrated in Figure 1. Most of our features are designed in such a way that they assign a ranking score to each of our food items by counting how often a feature is observed with a food item; that is why we call these features **ranking features**. The resulting ranking should assign high scores to food brands and low scores to food types. If we want to combine several features into one ranking, we simply average for each food item the different ranking scores of the individual ranking features. This is possible since they have the same range $[0; 1]$. We obtain such range by normalizing the number of occurrences of a feature with a particular food item by the total number of occurrences of that food item. The combination by averaging is unbiased as it treats all features equally.

We also introduce a **reset feature** which is applied on top of an existing ranking provided by ranking features. A reset feature is a negative feature in the sense that it is usually a reliable cue that a food item is *not* a brand. If it fires for a particular food item, then its ranking score is reset to 0.

Finally, we add **bootstrapping features**. These features produce an output similar to the ranking features (i.e. another ranking). However, unlike the ranking features, the bootstrapping features produce their output based on a weakly-supervised method which requires some labeled input. Rather than manually providing that input, we derive it from the combined output that is provided by the ranking and reset features. We restrict ourselves to instances with a high-confidence prediction, which translates to the top and bottom end of a ranking. (Since the instances are not manually labeled, of course, not every label assignment will be correct. We hope, however, that by restricting to instances with a high-confidence prediction, we can reduce the amount of errors to a minimum.) The output of a bootstrapping feature is combined with the set of ranking features to a new ranking onto which again a reset feature is applied.

Table 4 shows which feature (each will be discussed below) belongs to which of the above feature types (i.e. ranking, reset or bootstrapping features). Most features (i.e. all except WIKI) are extracted from our domain-specific corpus introduced in §2.

4.1 Length

Since we established that brands tend to be shorter than types (§3), we add one feature that ranks each food item according to its number of characters.

4.2 Target Named-Entity Recognition (NER$_{target}$)

Brands can be considered a special kind of named entities. We apply a part-of-speech tagger to count how often a food item has been tagged as a proper noun. We decided against a named-entity recognizer as it usually only recognizes persons, locations and organizations, while part-of-speech taggers employ a general tag for all proper nouns (that may go well beyond the three afore-mentioned common types). We use a statistical tagger, i.e. *TreeTagger* (Schmid, 1994), that also employs features below the word level. As many of our food items will be unknown words, a character-level analysis may still be able to make useful predictions.

4.3 Contextual Named-Entity Recognition (NER$_{context}$)

We also count the number of other named entities that co-occur with the target food brand within the same sentence. We are only interested in organizations; an organization co-occurring with a brand is likely to be the company producing that brand (e.g. *He loves Kellogg's*$_{company}$ *frosties*$_{brand}$.) For this feature, we rely on the output of a named-entity recognizer for German (Chrupała and Klakow, 2010).

4.4 Diversification (DIVERS)

Once a product has established itself on the market for a substantial amount of time, many companies introduce variants of their brand to further consolidate their market position. The purpose of this diversification is to appeal to customers with special needs. A typical variant of food brands are *light* products. In many cases, the names of variants consist of the name of the original brand with some prefix or suffix indicating the particular type of variant (e.g. *mini babybel* or *philadelphia light*). We manually compiled 11 affixes and check for each food item how often it is accompanied by one of them.

4.5 Commerce Cues (COMMERCE)

Presumably, brands are more likely to be mentioned in the context of commercial transaction events than types. Therefore, we created a list of words that indicate these types of events. The list was created ad hoc. We used external resources, such as FrameNet (Baker et al., 1998) or GermaNet (Hamp and Feldweg, 1997) (the German version of WordNet (Miller et al., 1990)), and made no attempt to tune that list to our domain-specific food corpus. The final list (85 cues in total) comprises: verbs (and deverbal nouns) that convey the event of a commercial transaction (e.g. *buy*, *purchase* or *sell*), persons involved in a commercial transaction (e.g. *customer* or *shop assistant*), means of purchase (e.g. *money*, *credit card* or *bill*), places of purchase (e.g. *supermarket* or *shop*) and judgment of price (e.g. *cheap* or *expensive*).

4.6 Food Modifier (PAT$_{mod}$)

Even though many mentions of brands are similar to those of types, there exist some particular contexts that are mostly observed with brands. If the food item to be classified often occurs as a modifier of another food item, then the target item is likely to be some brand. This is due to the fact that many brands are often mentioned in combination with the food type that they represent, e.g. ***volvic*** *mineral water*, ***nutella*** *chocolate spread*.

4.7 Prepositional Phrase Embedding (PAT$_{pp}$)

Instead of appearing as a modifier (§4.6), a brand may also be embedded in some prepositional phrase that has a similar meaning, e.g. *We only buy the chocolate spread* $[$*by* ***nutella***$]_{PP}$.

4.8 Graph-based Methods (GRAPH)

We also employ some semi-supervised graph clustering method in order to assign semantic types to food items as introduced in Wiegand et al. (2014). The underlying data structure is a food graph that is generated automatically from our domain-specific corpus where nodes represent food items and edge weights

Category	Description	General	Brands
MEAT	meat and fish (products)	19.48	1.31
BEVERAGE	beverages (incl. alcoholic drinks)	17.19	**23.96**
SWEET	sweets, pastries and snack mixes	14.90	**25.60**
SPICE	spices and sauces	10.53	2.42
VEGE	vegetables (incl. salads)	10.38	0.00
STARCH	starch-based side dishes	9.21	4.42
MILK	milk products	6.71	**23.48**
FRUIT	fruits	4.48	1.14
GRAIN	grains, nuts and seeds	3.41	0.00
FAT	fat	2.54	**20.00**
EGG	eggs	0.92	0.00

Table 5: Proportion of categories in the entire food vocabulary (*General*) and among brands (*Brands*).

represent the similarity between different items. The weights are computed based on the frequency of co-occurrence within a similarity pattern (e.g. *X instead of Y*). Food items that cluster with each other in such a graph (i.e. food items that often co-occur in a similarity pattern) are most likely to belong to the same class. For the detection of brands, we examine two different types of food categorization. We always use the same clustering method (Wiegand et al., 2014) and the same graph. Depending on the specific type of categorization, we only change the seeds to fit the categories to be induced.

4.8.1 Categories of the Food Guide Pyramid (GRAPH$_{pyramid}$)

The first categorization we consider is the categorization of food items according to the *Food Guide Pyramid* (U.S. Department of Agriculture, 1992) as examined in Wiegand et al. (2014). We observed that food brands are not equally distributed throughout the entire range of food items. There is a notable bias of food brands towards beverages (mostly soft drinks and alcoholic drinks), sweets, snack mixes, dairy products and fat. Other categories, e.g. nuts, vegetables or meat, hardly contain brands.[4] The category inventory and the proportion among types and brands are displayed in Table 5.

We use the category information as a negative feature, that is, we re-set the ranking score to 0 if the category of the food item is either MEAT, SPICE, VEGE, STARCH, FRUIT, GRAIN or EGG. In order to obtain a category assignment to our food vocabulary, we re-run the best configuration from Wiegand et al. (2014) including the choice of category seeds. We just extend the graph that formerly only contained food types by nodes representing brands. We use no manually-compiled knowledge regarding food brands. Even though the seed food items are exclusively food types, we hope to be also able to make inferences regarding food brands. This is illustrated in Figure 2(a): The brand *mars* can be grouped with food types that are sweets, therefore, we conclude that *mars* is also some sweet. (Brands can be grouped with food types of their food category, since food brands are often used as if they were types (§1)). Since sweets are plausible candidates for brands (Table 5), *mars* is likely to be some brand.

We think that such bias of brands towards certain subcategories is also present in other domains. For example, in the electronic domain laptops will have a much larger variety of brands than network cables. Similarly, in the fashion domain there exist much more shoe brands than sock brands.

4.8.2 Direct Graph Clustering Separating Brands from Types (GRAPH$_{brand}$)

We also apply graph clustering directly for the separation of brands from types, i.e. we assign some brand and type seeds and then run graph-based clustering (Figure 2(b)). In order to combine the output of this clustering with that of the previous methods, we interpret the confidence of the output as a ranking score. As we pursue an unsupervised approach, we do not manually label the seeds but rely on the output of a ranker using a combination of above features (Figure 1). Instances at the top of the ranking are considered brand seeds, while instances at the bottom are considered type seeds.

[4]There may be companies which, among other things, also sell these food types, but we do not want to extract the names of *organizations* (as in traditional named-entity recognition), e.g. *Kraft Foods*, but specific product names, e.g. *philadelphia*.

(a) food type categorization (b) brand detection

Figure 2: Similarity graphs; **bold** items are seeds; line width of edges represents strength of similarity.

4.9 Wikipedia Bootstrapping (WIKI)

For many information extraction tasks, the usage of collaboratively-edited resources is increasingly becoming popular. One of the largest resources of that type is *Wikipedia*. For our vocabulary of food items, we could match 57% of the food brands and 53% of the food types with a Wikipedia article.

Even though Wikipedia may hold some useful information for the detection of brands, this information is not readily available in a structured format, such as *infoboxes*. This is illustrated by (3)-(5) which display the first sentence of three Wikipedia articles, where (3) and (4) are food brands and (5) is a food type. There is some thematic overlap across the two categories (e.g. (4) and (5) describe the ingredients of the food item). However, if one also considers the entire articles, some notable topical differences between brands and types become obvious. The articles of food brands typically focus on commercial aspects (i.e. market situation and product history) while articles of food types describe the actual food item (e.g. by distinguishing it from other food items or naming its origin). Therefore, a binary topic classification based on the *entire* document should be a suitable approach. In the light of the diversified language employed for articles on brands (cp. (3)-(4)), we consider a bag-of-words classifier more effective than applying some textual patterns on those texts.

(3) *BRAND:* **Twix** is a chocolate bar made by Mars, Inc.

(4) *BRAND:* **Smarties** is a brand under which Nestlé produces colour-varied sugar-coated chocolate lentils.

(5) *TYPE:* **Milk chocolate** is a type of chocolate made from cocoa produce (cocoa bean, cocoa butter), sugar, milk or dairy products.

Similar to GRAPH$_{brand}$ (§4.8.2), we harness Wikipedia via a bootstrapping method. We generate a labeled training set of Wikipedia articles representing brands and types using the combined output of the ranking features (+ reset feature). We then train a supervised classifier on these data and classify all articles representing food items of our food vocabulary. We use the output score of the classifier for the article of each food item (which amounts to some confidence score) and thus obtain a ranking score. For those food items for which no Wikipedia entry exists, we produce a score of 0.

4.10 Vector Space Model (VSM)

While GRAPH$_{brand}$ (§4.8.2) determines similar food items by means of highly weighted edges in a similarity graph (that represent the frequency of co-occurrences with a similarity pattern), we also examine whether distributional similarity can be harnessed for the same purpose. We represent each food item as a vector, where the vector components encode the frequency of words that co-occur with mentions of the food item in a fixed window of 5 words (in our domain-specific corpus). Similar to GRAPH$_{brand}$ (§4.8.2) and WIKI (§4.9), we consider the n highest and m lowest ranked food items provided by ranking features (+ reset feature) as labeled brand and type instances for a supervised classifier. For testing, we apply this classifier on each food item in our vocabulary, or more precisely, its vector representation. Thus we obtain another ranking score (again, the output amounts to some confidence score).

	Plain					+Graph$_{pyramid}$ *(reset feature)*				
Feature	**P@10**	**P@50**	**P@100**	**P@200**	**AP**	**P@10**	**P@50**	**P@100**	**P@200**	**AP**
RANDOM	10.00	18.00	14.00	14.00	0.119	20.00	22.00	22.00	21.50	0.167
LENGTH	10.00	20.00	22.00	21.50	0.163	10.00	32.00	41.00	40.00	0.230
DIVERS	60.00	46.00	37.00	25.00	0.207	60.00	50.00	39.00	30.50	0.240
COMMERCE	30.00	28.00	31.00	27.00	0.220	40.00	38.00	39.00	35.00	0.294
NER$_{context}$	70.00	72.00	52.00	43.50	0.401	80.00	72.00	51.00	46.50	0.425
PAT$_{pp}$	90.00	78.00	64.00	50.00	0.439	**100.00**	78.00	69.00	53.00	0.476
PAT$_{mod}$	60.00	68.00	**69.00**	58.00	0.460	90.00	76.00	**76.00**	58.00	0.507
NER$_{target}$	80.00	70.00	60.00	52.50	0.479	80.00	78.00	72.00	61.50	0.525
combined	**100.00**	**88.00**	66.00	**59.00**	**0.612**	**100.00**	**86.00**	**76.00**	**62.50**	**0.626**

Table 6: Precision at rank n (P@n) and average precision (AP) of the different ranking features.

Partition	**Prec**	**Rec**	**F**
Food Types	70.49	72.82	71.04
Food Brands	69.09	66.21	64.93

Table 7: Performance of food categorization according to the *Food Guide Pyramid* (auxiliary classification).

5 Experiments

In the following experiments, we mostly evaluate rankings. For that we employ *precision at rank n* and *average precision*. The former computes precision at a predefined rank n, whereas the latter provides an average of the precisions measured at every possible rank. While average precision provides a score that evaluates the ranking as a whole, precision at rank n typically focuses on the correctness of higher ranks.[5]

5.1 Evaluation of Ranking Features

Table 6 (left half) displays the results of the individual and combined ranking features. As a trivial baseline, we also include RANDOM which is randomized ranking of the food items. The table shows that all features except LENGTH produce a notably better ranking than RANDOM. Following the inspection of intrinsic properties of brands in §3, it does not come as a surprise that NER$_{target}$ is the strongest feature. However, also the contextual features NER$_{context}$, PAT$_{pp}$ and PAT$_{mod}$ produce reasonable results. If we combine all features (except the poorly performing LENGTH), we obtain a notable improvement over NER$_{target}$ which proves that those different features are complementary to a certain extent.

5.2 Evaluation of the Reset Feature

In Table 7, we examine the food categorization according to the Food Guide Pyramid as such. For this evaluation, we partition the output of automatic categorization into (actual) types and brands. Thus we can compare the performance between those two different types of food items, and can quantify the loss on the categorization on brands against the categorization on types. (Due to the fact that the seeds exclusively comprise types, we must assume that performance on brands will be lower.)[6] Even though there is a slight loss on brands (mostly recall), we still consider this categorization useful for our purposes.

[5]The manually labeled food vocabulary is available at:
`www.lsv.uni-saarland.de/personalPages/michael/relFood.html`

[6]Since the categories to indicate unlikely brands (§4.8.1) are extremely sparse (Table 5), we conflate them for this evaluation as one large category NEGATIVE. Because of this and due to the fact that the food type vocabulary is slightly smaller than the one used in Wiegand et al. (2014) (since we only consider food items mentioned at least 5 times in our corpus (§2)), the performance scores of food categorization in Table 7 and the one reported in Wiegand et al. (2014) differ accordingly.

Classifier		Acc	Prec	Rec	F
Baselines	Majority-Class Classifier	88.76	44.38	50.00	47.02
	seeds only: 50 top+150 bottom	9.51	91.00	13.85	23.47
	seeds only: 100 top+300 bottom	18.57	86.17	25.48	37.81
Bootstrap. Features	WIKI (*seeds: 50 top+150 bottom*)	43.95	**87.68**	43.33	57.91
	VSM (*seeds: 100 top+300 bottom*)	77.87	64.93	**81.61**	66.39
	GRAPH$_{brand}$ (*seeds: 100 top+300 bottom*)	82.91	81.36	67.27	**73.53**

Table 8: Bootstrapping features in isolation compared with baselines (i.e. reference classifiers).

Table 6 (right half) shows the performance of the corresponding reset feature on the brand detection task. We observe a systematic increase in performance when added on top of the ranking features.

5.3 Evaluation of Bootstrapping Features

Table 9 displays the performance of the bootstrapping features. For the labeled training data, we empirically determined the optimal class ratio (1:3) and the optimal number of seeds (the top 100 and bottom 300 items for VSM and GRAPH$_{brand}$, and top 50 and bottom 150 items for WIKI). As a supervised classifier for VSM and WIKI, we chose Support Vector Machines using *SVM^{light}* (Joachims, 1999).

The table shows that only GRAPH$_{brand}$ and WIKI improve the ranking, whereas WIKI is notably stronger. These results suggest that Wikipedia is a good resource from which to learn whether a food item is a brand or not. However, this task could not be completely solved by WIKI since not all food items are covered by Wikipedia (§4.9). To further prove this, we also evaluate an upper bound of Wikipedia, WIKI$_{oracle}$ (exclusively using that resource), in which we pretend to correctly interpret every Wikipedia page as an article for either a food brand or a food type. We rank all brands having a Wikipedia article highest. They are followed by those food items having no article (ordered randomly) and, finally, by the food types having a Wikipedia article. Table 9 shows that we are able to outperform WIKI$_{oracle}$.

Our pipeline (Figure 1) applies the reset feature at two stages. We also examine whether it is necessary to apply that feature for a second time. Presumably, the bootstrapping feature is so effective that we do not have to apply further type filtering. After all, the reset feature will also downweight some correct food items (Table 5). Table 9 confirms that when the reset feature is applied only once, we obtain a better performance (according to average precision) for all bootstrapping features (even for VSM).

Finally, Table 8 evaluates the bootstrapping features in isolation. Since, unlike the ranking features, the bootstrapping features provide a definite classification for each food item (in addition to a prediction score evaluated as a ranking score), we consider the output for a binary classification task. In this setting, we make use of the four evaluation measures *accuracy*, *precision*, *recall* and *F-score*. For the last three measures, we always compute the macro average score.

As a baseline, we also include a majority-class classifier that always predicts the class *food type*. Interestingly, in terms of F-score, GRAPH$_{brand}$ is the best method rather than WIKI, i.e. the best method from the previous evaluation in Table 9. The reason for this is that we evaluate in isolation rather than in combination with other features (i.e. parts of the additional benefit included in GRAPH$_{brand}$ may already be contained in ranking and reset features). Secondly, in a ranking task (Table 9), good performance is usually achieved by classifiers biased towards a high precision. Indeed, the best ranker in Table 9, i.e. WIKI, achieves the highest precision in Table 8.

6 Related Work

Ling and Weld (2012) examine named-entity recognition on data that also include brands, however, the class of brands is not explicitly discussed. Putthividhya and Hu (2011) explore brands in the context of product attribute extraction. Entities are extracted from eBay's clothing and shoe category. Nadeau et al. (2006) explicitly generate gazetteers of car brands obtained from corresponding websites. Those textual data are very restrictive in that they do not represent sentences but category listings or tables. In this paper, we consider as textual source a more general text type, i.e. forum entries, that comprise full

			-2nd reset	
Feature	**P@200**	**AP**	**P@200**	**AP**
WIKI$_{oracle}$	66.00	0.429	*-N/A-*	*-N/A-*
ranking+GRAPH$_{pyramid}$	62.50	0.626	*-N/A-*	*-N/A-*
ranking+GRAPH$_{pyramid}$+VSM	60.00	0.619	63.00	0.661
ranking+GRAPH$_{pyramid}$+GRAPH$_{brand}$	67.50	0.638	65.50	0.662
ranking+GRAPH$_{pyramid}$+WIKI	**70.00**	**0.688**	**73.00**	**0.718**

Table 9: Impact of bootstrapping; *-2nd reset:* does not apply reset feature for a second time (Figure 1).

sentences. Previous work also focuses on traditional (semi-)supervised algorithms. Hence, there are only few additional insights as to the specific properties of brand names. Min and Park (2012) examine the aspect of product instance distinction on the use case of product reviews on jeans from Amazon. Their work focuses on temporal features to identify distinct product instances (these may also include brand names).

The food domain has also recently received some attention. Different types of classification have been explored including ontology mapping (van Hage et al., 2005), part-whole relations (van Hage et al., 2006), recipe attributes (Druck, 2013), dish detection and the categorization of food types according to the Food Guide Pyramid (Wiegand et al., 2014). Relation extraction tasks have also been examined. While a strong focus is on food-health relations (Yang et al., 2011; Miao et al., 2012; Kang et al., 2013; Wiegand and Klakow, 2013), relations relevant to customer advice have also been addressed (Wiegand et al., 2012; Wiegand et al., 2014). Beyond that, Chahuneau et al. (2012) relate sentiment information to food prices with the help of a large corpus consisting of restaurant menus and reviews. Druck and Pang (2012) extract actionable recipe refinements. To the best of our knowledge, we present the first work that explicitly addresses the detection of brands in the food domain. While brands as such present an additional dimension to previously examined types of categorization, we also show that the categorization according to the Food Guide Pyramid helps to decide whether a food item is a brand or not.

7 Conclusion

We examined the task of separating types from brands in the food domain. Framing the problem as a ranking task, we directly converted predictive features extracted from a domain-specific corpus into a ranker without the need of labeled training data. Apart from those ranking features, we also exploited knowledge induced by semi-supervised graph-based clustering for two different purposes. On the one hand, we produced an auxiliary categorization of food items according to the Food Guide Pyramid, and assumed that a food item is a type when it belongs to a category that is unlikely to contain brands. On the other hand, we directly modelled the task of brand detection by using seeds provided by the output of the textual ranking features. We also learned additional high-precision knowledge from Wikipedia webpages using a similar bootstrapping scheme.

Acknowledgements

This work was performed in the context of the Software-Cluster project SINNODIUM. Michael Wiegand was funded by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IC10S01. The authors would like to thank Melanie Reiplinger for proofreading the paper.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pages 86–90, Montréal, Quebec, Canada.

Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis, and Noah A. Smith. 2012. Word Salad: Relating Food Prices and Descriptions. In *Proceedings of the Joint Conference on Empirical Methods in Natural*

Language Processing and Computational Natural Language Learning (EMNLP/CoNLL), pages 1357–1367, Jeju Island, Korea.

Grzegorz Chrupała and Dietrich Klakow. 2010. A Named Entity Labeler for German: Exploiting Wikipedia and Distributional Clusters. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 552–556, La Valletta, Malta.

Gregory Druck and Bo Pang. 2012. Spice it up? Mining Refinements to Online Instructions from User Generated Content. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 545–553, Jeju, Republic of Korea.

Gregory Druck. 2013. Recipe Attribute Detection Using Review Text as Supervision. In *Proceedings of the IJCAI-Workshop on Cooking with Computers (CWC)*, Beijing, China.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 174–181, Madrid, Spain.

Thorsten Joachims. 1999. Making Large-Scale SVM Learning Practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press.

Jun Seok Kang, Polina Kuznetsova, Michael Luca, and Yejin Choi. 2013. Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1443–1448, Seattle, WA, USA.

Richard R. Klink. 2000. Creating Brand Names with Meaning: The Use of Sound Symbolism. *Marketing Letters*, 11(1):5–20.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press.

Xiao Ling and Daniel S. Weld. 2012. Fine-Grained Entity Recognition. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 94–100, Toronto, Canada.

Qingliang Miao, Shu Zhang, Bo Zhang, Yao Meng, and Hao Yu. 2012. Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text. In *Proceedings of the Pacific Asia Conference on Language, Information and Compuation (PACLIC)*, pages 99–107, Bali, Indonesia.

George Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3:235–244.

Hye-Jin Min and Jong C. Park. 2012. Product Name Classification for Product Instance Distinction. In *Proceedings of the Pacific Asia Conference on Language, Information and Compuation (PACLIC)*, pages 289–298, Bali, Indonesia.

David Nadeau, Peter D. Turney, and Stan Matwin. 2006. Unsupervised Named-Entity Recognition: Generating Gazetteers and Resolving Ambiguity. In *Proceedings of the Canadian Conference on Artificial Intelligence*, pages 266–277, Québec City, Québec, Canada.

Duangmanee Putthividhya and Junling Hu. 2011. Bootstrapped Named Entity Recognition for Product Attribute Extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1557–1567, Edinburgh, Scotland, UK.

Ellen Riloff and Jessica Shepherd. 1997. A Corpus-Based Approach for Building Semantic Lexicons. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 117–124, Providence, RI, USA.

Brian Roark and Eugene Charniak. 1998. Noun-phrase co-occurrence statistics for semi-automatic semantic lexicon construction. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 1110–1116, Montreal, Quebec, Canada.

Kim Robertson. 1989. Strategically Desirable Brand Name Characteristics. *Journal of Comsumer Marketing*, 6(4):61–71.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49, Manchester, United Kingdom.

Human Nutrition Information Service U.S. Department of Agriculture. 1992. The Food Guide Pyramid. Home and Garden Bulletin 252, Washington, D.C., USA.

Willem Robert van Hage, Sophia Katrenko, and Guus Schreiber. 2005. A Method to Combine Linguistic Ontology-Mapping Techniques. In *Proceedings of International Semantic Web Conference (ISWC)*, pages 732 – 744, Galway, Ireland. Springer.

Willem Robert van Hage, Hap Kolb, and Guus Schreiber. 2006. A Method for Learning Part-Whole Relations. In *Proceedings of International Semantic Web Conference (ISWC)*, pages 723 – 735, Athens, GA, USA. Springer.

Michael Wiegand and Dietrich Klakow. 2013. Towards Contextual Healthiness Classification of Food Items – A Linguistic Approach. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, pages 19–27, Nagoya, Japan.

Michael Wiegand, Benjamin Roth, and Dietrich Klakow. 2012. Web-based Relation Extraction for the Food Domain. In *Proceedings of the International Conference on Applications of Natural Language Processing to Information Systems (NLDB)*, pages 222–227, Groningen, the Netherlands. Springer.

Michael Wiegand, Benjamin Roth, and Dietrich Klakow. 2014. Automatic Food Categorization from Large Unlabeled Corpora and Its Impact on Relation Extraction. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 673–682, Gothenburg, Sweden.

Hui Yang, Rajesh Swaminathan, Abhishek Sharma, Vilas Ketkar, and Jason D'Silva, 2011. *Learning Structure and Schemas from Documents*, volume 375 of *Studies in Computational Intelligence*, chapter Mining Biomedical Text Towards Building a Quantitative Food-disease-gene Network, pages 205–225. Springer Berlin Heidelberg.

Eric Yorkston and Geeta Menon. 2004. A Sound Idea: Phonetic Effects of Brand Names on Consumer Judgments. *Journal of Consumer Research*, 31:43–51.

Patrick Ziering, Lonneke van der Plas, and Hinrich Schuetze. 2013. Bootstrapping Semantic Lexicons for Technical Domains. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, Nagoya, Japan, 1321–1329.