Implicit Linking of Food Entities in Social Media

Wen-Haw Chong and Ee-Peng Lim

Singapore Management University, 80 Stamford Road, Singapore 178902,
whchong.2013@phdis.smu.edu.sg, eplim@smu.edu.sg

Abstract. Dining is an important part in people's lives and this explains why food-related microblogs and reviews are popular in social media. Identifying food entities in food-related posts is important to food lover profiling and food (or restaurant) recommendations. In this work, we conduct Implicit Entity Linking (IEL) to link food-related posts to food entities in a knowledge base. In IEL, we link posts even if they do not contain explicit entity mentions. We first show empirically that food venues are *entity-focused* and associated with a limited number of food entities each. Hence same-venue posts are likely to share common food entities. Drawing from these findings, we propose an IEL model which incorporates venue-based query expansion of test posts and venue-based prior distributions over entities. In addition, our model assigns larger weights to words that are more indicative of entities. Our experiments on Instagram captions and food reviews shows our proposed model to outperform competitive baselines.

Keywords: entity linking, food entities, query expansion

1 Introduction

In social media, food-related topics are highly popular. Many users post food-related microblogs or reviews on various platforms such as Instagram, Foursquare, Yelp, etc. Such user generated content can be mined for profiling food lovers or for food and dining venue recommendations. In fact, identifying the local cuisines in posts has been justified [13] as useful for helping tourists in their dining choices. In this work, we propose to link food-related posts to a knowledge base of food entities. Given a test post that mention or *merely imply* some food entity, the task is to rank food entities in order of relevance.

We refer to this problem of linking posts as *Implicit Entity Linking* (IEL) [14, 11]. In IEL, one links each test post to one or more related entities, without the need for mention extraction. This contrasts with the Explicit Entity Linking (EL) problem [10, 19, 9, 18, 17] which links mentions of named entities. Notably IEL circumvents the challenge of mention extraction in social media where posts are often grammatically noisy and colloquial. IEL also generalizes easily to various content scenarios. For example, consider the text snippets "XX Chicken Rice", "rice with chicken" and "having lunch". These are cases where food entities are respectively mentioned via proper nouns, improper nouns and merely implied.

All snippets can be processed via IEL while EL is mention-dependent and will process only the first snippet comprising proper nouns. Lastly, IEL is also easier to conduct if one is only focused on a certain entity type, e.g. food entities. There is no need to ensure that only mentions of the right type are extracted.

Problem Setting. We formulate IEL as a ranking problem. For each post, we rank candidate food entities such that high ranking entities are more likely to be related. We assume that posts are not labeled with food entities for training, but are associated with posting venues. Both assumptions are realistic. Firstly labeled data are typically expensive to obtain. Secondly venue information is often available for platforms such as Foursquare, Instagram, review websites etc. We use Wikipedia as the knowledge base to link against.

Contributions. Our contributions are (1) an empirical analysis whereby we highlight that venues are focused around a limited set of food entities each, i.e., *entity-focused characteristic* and (2) a series of models for IEL. Our best performing model comprises the following aspects:

- **Entity-Indicative Weighting**: We propose a weighting scheme in our model to assign more weights to entity-indicative words. The intuition is that such words are more important for inferring entities than other words.
- **Query Expansion:** The entity-focused characteristic implies that a test post is likely to share common food entities as other same-venue posts. Hence we augment each test post via query expansion to include words from other same-venue posts.
- **Venue-based Prior:** Leveraging the same entity-focused characteristic, we generate venue-based prior distribution over food entities in an initial entity linking stage. This prior is used to bias the entity scores for the next stage.

By combining all above aspects, our best model EW-EWQE(v) outperforms state-of-the-art baselines that have been adapted for implicit entity linking.

2 Empirical Analysis

2.1 Datasets

In our empirical analysis and subsequent experiments, we use data from Instagram and Burpple [1]. The latter is a popular food review website in Singapore. Both datasets are generated by users from Singapore, a city well known for its wide range of food choices. Since both datasets are from Singapore users, we link their posts against a list of 76 food entities derived from the Wikipedia page on Singapore's cuisines[2]. Further details are discussed in Section 4.1.

For Instagram, we collect highly popular food-related captions from 2015 using hashtags of food e.g. '#foodporn' [3], or food entities e.g. '#chillicrab'. Following data cleaning and venue deduplication, we obtained 278,647 Instagram

[1] https://www.burpple.com/sg

[2] https://en.wikipedia.org/wiki/Singaporean_cuisine

[3] the most popular food related hashtag on our Instagram dataset

posts from 79,496 distinct venues. For Burpple, all its posts are food reviews and filtering by hashtags is not required. From Burpple, we obtained 297,179 posts over 13,966 venues. Table 1 illustrates four sample posts, two each from Instagram and Burpple. Clearly some posts are more informative about food entities than others. For example, the first instagram example does not reveal the food entity explicitly while the second example mentions fish ball noodle.

Table 1. Sample posts comprising Instagram captions and Burpple reviews.

Instagram	“super heavy lunch. and spicy! but its a must-try cafe! #food #foodporn #foodie #foodgasm #badoquecafe #instagood”
	“yesterday’s lunch! #fishballnoodle #food #foodporn the soup was damn good”
Burpple	“*Signature Lamb Rack ($46++)* Very neat rectangular bricks of lamb, which we requested to be done medium-well.Nothing too impressive.. hurhur. Service is top -notch though”
	“*Good morning!* One of my favourite old school breakfast but he not his fav”

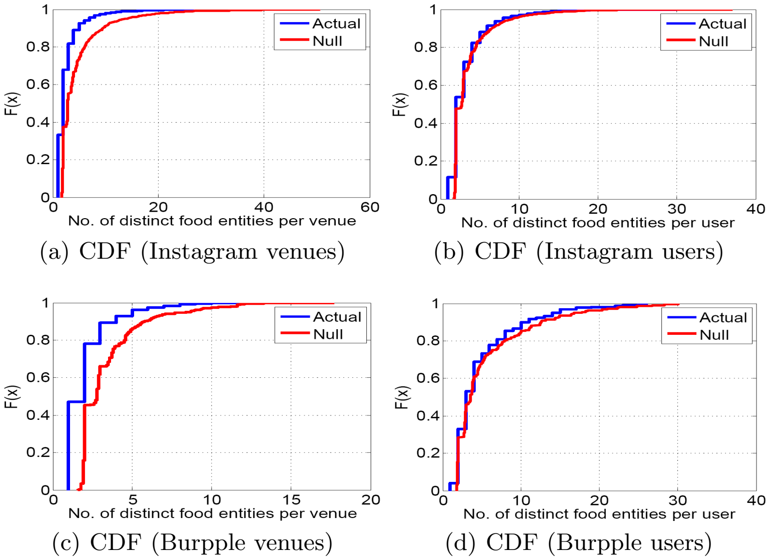
2.2 Analysis

A food venue typically focuses on some cuisines or food themes and is unlikely to serve an overly wide variety of dishes. For example, it is more probable for a restaurant to serve either Western or Asian cuisines, rather than both. Consequently, each food venue is likely to be associated with a limited number of food entities. We termed this as the *entity-focused characteristic*. To verify this characteristic, we compare the number of distinct food entities per venue against a null model where the characteristic is absent. We expect food venues to be associated with fewer food entities when compared against the null model.

For each venue v with multiple posts, we first compute the number of distinct entities over its posts. We then compute the expected number of distinct entities under the null model following the steps below:

- For each post from v, sample an entity e based on global entity probability i.e. entity popularity. Add to entity set $\mathbb{E}_{null}(v)$.
- Compute $|\mathbb{E}_{null}(v)|$, the distinct entity count under the null model.

We conduct our analysis on 2308 venues from Instagram and 362 venues from Burpple which have at least two user-labeled posts each. Such posts contain hashtags with food entity names, e.g. ‘#chillicrab’, ‘#naan’ etc. As sampling is required for the null model, we conduct 10 runs and take the average expected food entity count for each venue. For further analysis, we also repeat a similar procedure for users to compare their actual and expected food entity count. The intuition is that users may possess the entity-focused characteristic as well due to food preferences or constraints e.g vegetarian. The user statistics are computed over 2843 Instagram users and 218 Burpple users.



(a) CDF (Instagram venues) (b) CDF (Instagram users)

(c) CDF (Burpple venues) (d) CDF (Burpple users)

Fig. 1. CDFs of actual and expected distinct food entities for venues and users. F(x) on y-axis is probability of venues or users with $\leq x$ distinct food entities.

Figure 1 plots the Cumulative Distribution Function (CDF) of distinct food entities for venues and users on both Instagram and Burpple, whereby distinct entity counts are on a per venue or user basis. In each graph, the blue line represents the actual count while the red line is for counts from the null model (averaged over 10 runs). For Figures 1(a) and (c) venues are shown to be focused around specific food entities such that on average, each venue has fewer distinct food entities than expected under the null model. For example in Figure 1(a), around 98% of the Instagram venues are associated with 10 distinct food entities or less in the actual data. In contrast, the null model has a corresponding proportion of around 91%. A similar trend can be observed for Burpple venues as shown in Figure 1(c). Thus, the entity-focused characteristic is clearly evident for the venues of both datasets.

Figures 1(b) and (d) plot for Instagram and Burpple users respectively. There is much less difference between the actual and null model count, as both the blue and red lines overlap substantially in both figures. Comparing the plots for venues and users, we conclude that users are relatively less focused on food entities when compared to venues. These findings have implications for entity linking and should be considered when designing models. In particular, given a test post with both user and venue information, it may be easier to improve linking accuracy by exploiting other posts from the same venue rather than from the same user. In Section 3.2, we shall introduce a query expansion approach based on exploiting the entity-focused characteristic of venues.

3 Models

In this section, we present a series of models for IEL, culminating in a final best performing model. We start with the naive Bayes model. This can be regarded as a standard information retrieval baseline. Let $\mathbf{w}$ be the set of words in a post, where for notation simplicity, we assume each unique word $w \in \mathbf{w}$ occurs only once in the post. In our problem setting, we assume the entity probability $p(e)$ to be uniform as labeled posts are unavailable for estimating $p(e)$. The probability of food entity e given $\mathbf{w}$ is:

$$p(e|\mathbf{w}) \propto \prod\nolimits_{w \in \mathbf{w}} p(w|e) = \prod\nolimits_{w \in \mathbf{w}} \frac{f(e,w) + \gamma}{\sum_{w'} f(e,w') + W\gamma} \tag{1}$$

whereby $f(e,w)$ is the number of co-occurrences of word w with entity e, γ is the smoothing parameter and W is the vocabulary size. In the absence of labeled posts, the co-occurrences are estimated solely from the Wikipedia knowledge base. For each food entity e, we derive $f(e,w)$ by the count of w occurrences in the Wikipedia page of e and in Wikipedia text snippets around hyperlinks to e (refer Section 4.1). Finally entities are ranked by $p(e|\mathbf{w})$. The naive Bayes model is efficient and highly amenable to extensions.

3.1 Entity-Indicative Weighting (EW)

The naive Bayes model multiplies word probabilities without considering which words are more important for entity linking. Intuitively, some words are more indicative of food entities than others and should be assigned greater importance in entity linking models. Formally, an entity-indicative word w has relatively high $p(e|w)$ for some entity/entities in comparison with other words, e.g. ‘sushi’ is more entity-indicative than ‘dinner’.

An entity-indicative word is different from a high probability word given an entity. For example, a food entity e may have high probability of generating the word ‘rice’, i.e. $p(\text{‘rice’}|e)$ is high. However if many other food entities are also related to rice, then the word may not indicate e with high probability i.e. low $p(e|\text{‘rice’})$. If a post $\mathbf{w}$ mentions other more entity-indicative words, e.g. related to ingredients or cooking style, then such words should be assigned more importance when computing $p(e|\mathbf{w})$.

To capture the above intuition, we propose the entity-indicative weighting (EW) model. This assigns continuous weights to words and incorporates easily into the naive Bayes model. Let $\beta(w)$ be the *entity-indicative* weight for word w. This weight $\beta(w)$ is added as an exponent to the term $p(w|e)$ in Equation 1. By taking the log to avoid underflow errors, we obtain the EW model:

$$\ln p(e|\mathbf{w}) \propto \sum\nolimits_{w \in \mathbf{w}} \beta(w) \ln p(w|e) \tag{2}$$

Interestingly, Equation (2) is similar in form to the weighted naive Bayes model proposed in prior work [22, 7] for classification tasks. Here, we use it for IEL.

To compute the weights $\beta(w)$, we apply the vector space model and treat entities as documents. By definition, entity-indicative words are associated with fewer entities and have large weights. Weights are defined by:

$$\beta(w) = \log(1 + E/df(w)) \tag{3}$$

where E is the number of distinct food entities considered and $df(w)$ counts number of food entities with at least one occurrence of w.

3.2 Query Expansion with Same-Venue Posts

Based on the entity-focused characteristic, we expect that as a venue accumulates posts over time, its set of entities will be discussed repeatedly over different posts. This implies that for a test post discussing some entity e, there may exist other same-venue posts related to e. Hence if we augment the test post appropriately with words from other same-venue posts, we can potentially overcome information sparsity in one post and improve entity linking accuracy. This strategy is also known as query expansion.

Let test post $\mathbf{w}$ be posted from venue v. The idea is then to score candidate words w' appearing in other posts from v and whereby $w' \notin \mathbf{w}$. The expanded words w's aim to provide additional information for inferring the latent entity in $\mathbf{w}$. Among the many scoring schemes in the literature, we adopt a relatively simple cosine similarity scheme from [4]. This scheme scores each candidate word w' by its average relatedness $0 \leq \alpha_v(w', \mathbf{w}) \leq 1$ to the test post as:

$$\alpha_v(w', \mathbf{w}) = \frac{1}{|\mathbf{w}|} \sum\nolimits_{w \in \mathbf{w}} \frac{d_v(w', w)}{\sqrt{d_v(w') d_v(w)}} \tag{4}$$

where $|\mathbf{w}|$ is the number of words in $\mathbf{w}$, $d_v(w', w)$ is the count of v's posts containing both w' and w; and $d_v(w)$ is the count of v's posts with w. Intuitively, if w' co-occurs more with each word from $\mathbf{w}$ on average, then average relatedness is higher. However, relatedness can be over-estimated for common words. To mitigate this, Equation (4) includes in the denominator the product of word frequencies as the normalization term.

Following query expansion using same-venue posts, we combine two different word sets in a weighted naive Bayes model, which we refer to as QE(v):

$$\ln p(e | \{\mathbf{w}, \mathbf{w'}\}, v) \propto \sum\nolimits_{w \in \mathbf{w}} \ln p(w|e) + \sum\nolimits_{w' \in \mathbf{w'}} \alpha_v(w', \mathbf{w}) \ln p(w'|e) \tag{5}$$

where $\mathbf{w'}$ is the set of added words for post $\mathbf{w}$ from venue v. Since $0 \leq \alpha_v(w', \mathbf{w}) \leq 1$, Equation (5) illustrates that the original query words $w \in \mathbf{w}$ have greatest importance in the model while the importance of newly added words $w' \in \mathbf{w'}$ vary based on how related they are to the query.

In our experiments, we shall also compared against a model variant QE(u), which selects augmenting words from same-user posts. As conjectured in Section 2.2, this model may be less likely to improve linking accuracy.

3.3 Fused Model (EWQE)

We now combine the EW and QE(v) models to create a new fused model called EWQE. Intuitively, we consider a word as important only when it is both entity-indicative *and* highly related to the test post. For example, if a word is not indicative of any entities, then it is less useful for entity linking even if it is present in the test post or is a highly related word based on Equation (4) . On the other hand, a non-related word may be indicative of some entity which is unrelated to the test post, such that test post augmentation with it introduces noise and lowers accuracy.

To model the discussed conjunction logic, we multiply the weights from entity-indicative weighting and query expansion to obtain the model EWQE(v):

$$\ln p(e|\{\mathbf{w}, \mathbf{w'}\}, v) \propto \sum_{w \in \mathbf{w}} \beta(w) \ln p(w|e) + \sum_{w' \in \mathbf{w'}} \beta(w')\alpha_v(w', \mathbf{w}) \ln p(w'|e) \quad (6)$$

Alternatively, one can combine entity-indicative weighting with user-based query expansion. We denote such a model as EWQE(u) and include it for experiments.

3.4 Venue-based Prior

In our final model, we augment the probabilistic generative process in Equation (6) with a venue-based prior distribution over entities $p(e|v)$. Let joint probability $p(e, \{\mathbf{w}, \mathbf{w'}\}, v)$ be factorized as $p(v)p(e|v)p(\{\mathbf{w}, \mathbf{w'}\}|e)$. We now need to compute $p(e|v)$ while $p(\{\mathbf{w}, \mathbf{w'}\}|e)$ can be computed as before with the EWQE(v) model. Assuming uniform venue probability $p(v)$ and incorporating a weighting term η $(0 \leq \eta \leq 1)$, we have:

$$\begin{aligned}\ln p(e|\{\mathbf{w}, \mathbf{w'}\}, v) \propto\ & \eta \ln p(e|v) + \\ & (1-\eta)\left(\sum_{w \in \mathbf{w}} \beta(w) \ln p(w|e) + \sum_{w' \in \mathbf{w'}} \beta(w')\alpha_v(w', \mathbf{w}) \ln p(w'|e)\right) \quad (7)\end{aligned}$$

Basically $p(e|v)$ bias the entity score in a venue-specific manner, rather than a post-specific manner as prescribed by query expansion. Given a set of training posts labeled with food entities, $p(e|v)$ is computed trivially. However in our setting, we assume no labeled posts for training. Hence we compute Equation (7) in a 2-stage process as follows:

- Stage 1: With a desired IEL model, link the training posts. For each venue v, compute the aggregated entity scores $\tilde{p}(e|v)$, eg. if using the EW model, we compute $\tilde{p}(e|v) = \sum_{\mathbf{w} \in v} p(e|\mathbf{w})$. Normalize $\tilde{p}(e|v)$ to obtain $p(e|v)$.
- Stage 2: Combine $p(e|v)$ with the scores from the EWQE(v) model as detailed in Equation (7) to derive the final entity scores for ranking.

4 Experiments

4.1 Setup

Our experiment setup is weakly supervised. Training posts are assumed to be unlabeled with respect to food entities. These training posts are used only for query expansion and for computing the venue prior over entities, but not for computing the entity profile $p(w|e)$. The entity profile $p(w|e)$ and entity-indicative weights $\beta(w)$ are computed using only Wikipedia pages. However, we retain a small validation set of entity-labeled posts for tuning model parameters with respect to the ranking metrics. Also, all posts are associated with posting venues, regardless of whether they are in the training, test or validation set.

For discussion ease, denote posts with food entity hashtags e,g, '#chillicrab' as type A posts and post without such hashtags as type B posts. Type A posts are easily associated with Wikipedia food entities, which facilitates the construction of test and validation sets. Our datasets contain a mixture of both post types. For Instagram, we have 18,333 type A vs 216,881 type B posts[4] whereas for Burpple, we have 1944 type A vs 200,293 type B posts. We conduct 10 experiment runs for each dataset, whereby in each run, we *mask the food entity hashtags* of type A posts and randomly assign 50% of them to the training set, 20% to the validation set and 30% to the test set. The type B posts are all assigned to the training set. Lastly, most of our type A posts contain only one food entity hashtag each, hence we use such single-entity posts for evaluation in our test set.

Food Entities. We consider 76 food entities that are defined by Wikipedia as local cuisines of Singapore[5], as well as associated with distinct pages/descriptions. For each entity e, we construct its profile, i.e. $p(w|e)$ from its Wikipedia description page and Wikipedia text snippets with hyperlinks to e. For example, the Wikipedia page 'Pakistani_cuisine' contains many hyperlinks to the food entity 'Naan' [6]. When building the profile for 'Naan', we include the preceding and succeeding 10 words around each hyperlink.

Models to be Evaluated. We evaluate the following models:

- NB: The naive Bayes model from Equation (1).
- EW: Entity-indicative weighting as indicated in Equation (2).
- QE(v): Venue-based query expansion whereby each test post is augmented with words from other same-venue posts, as indicated in Equation (5).
- QE(u): User-based query expansion whereby each test post is augmented with words from other same-user posts.
- EWQE(v): Fusion of venue-based query expansion and entity-indicative weighting as shown in Equation (6).

[4] Filtering by vocabulary has been applied, hence the numbers sum to less than the total food-related posts in Section 2.1.

[5] https://en.wikipedia.org/wiki/Singaporean_cuisine

[6] oven-baked flatbread

- EWQE(u): Fusion of user-based query expansion and entity-indicative weighting.
- NB-EWQE(v): In stage 1, we compute $p(e|v)$ with the NB model, which is then combined with the EWQE(v) model in stage 2. See Equation (7).
- EW-EWQE(v): In stage 1, we use the EW model to compute $p(e|v)$. In stage 2, the computed $p(e|v)$ is combined with EWQE(v) model to score entities.

For each model, we use the validation set to tune γ, the smoothing parameter for $p(w|e)$, based on the grid [0.01, 0.1, 1, 10]. For NB-EWQE(v) and EW-EWQE(v), γ is jointly tuned with η whereby η is varied in steps of 0.1 from 0 to 1.

For further comparison, we introduce three other baselines. We adapt two EL models from [6, 5] such that they can be used for IEL. Without any adaptation, it is impossible for the vanilla EL models to link posts directly to entities. Our adaptations also aim to exploit the entity-focused characteristic of venues, or other related characteristics. Lastly, we include a word embedding baseline [20] that does not require any adaptation. The baselines are:

- TAGME: In the TAGME model [6, 15], candidate entities for a mention are voted for by candidate entities from other mentions in the same post. Adapting the idea to IEL, candidate entities for a post are voted for by candidate entities from other posts in the same venue. Since a candidate entity gathers larger votes from the same or related entities, this voting process exploits the entity-focused characteristic of venues as well. Let $\mathbf{w}_{i,v}$ denote the i-th post from venue v. Then candidate entity e_i for $\mathbf{w}_{i,v}$ gathers a vote from $\mathbf{w}_{j,v}$ computed as

$$vote(\mathbf{w}_{j,v} \rightarrow e_i) = \frac{1}{|e_j : p(e_j|\mathbf{w}_{j,v}) > 0|} \sum_{e_j : p(e_j|\mathbf{w}_{j,v})>0} sr(e_i, e_j) p(e_j|\mathbf{w}_{j,v}) \tag{8}$$

  where $sr(e_i, e_j)$ is the Jaccard similarity of incoming Wikipedia links [15] between e_i, e_j, and $p(e_j|\mathbf{w}_{j,v})$ can be based on any implicit entity linking models. Finally for ranking entities, we compute the final score for entity e_i as $p(e_i|\mathbf{w}_{i,v}) \sum_j vote(\mathbf{w}_{j,v} \rightarrow e_i)$.
- LOC: Locations in the form of grid cells can be entity-focused as well. To exploit this, we implement the framework from [5]. For each grid cell, the distributions over entities are inferred via EM learning and integrated with implicit entity linking models. Unlike [5], we omit the dependency on posting time as our targeted posts include food reviews which are usually posted after, rather than during meal events. We tune grid cell lengths based on grid [200m, 500m, 1km, 2km].
- PTE: This is a graph embedding method [20] that learns continuous vector representation for words, posts and entities over a heterogeneous graph. The graph consists of word nodes, post nodes and entity nodes, connected via the following edge types: word-word, post-word and entity-word. For each test post, we compute its vector representation by averaging over the representations of its constituent words. We then compute the cosine similarities to

entity representations for ranking. As in [20], we use an embedding dimension of 100. We set the number of negative samples to be 200 million.

For the baselines TAGME and LOC, we integrate the implicit entity linking models NB, EW and EW-EWQE(v). For each model, we replace the relevant mention-to-entity computations with post-to-entity computations. For example, TAGME(NB) computes $p(e_j|\mathbf{w}_{j,v})$ in Equation (8) using the NB model. Such integration leads to the following baseline variants: TAGME(NB), TAGME(EW), TAGME(EW-EWQE(v)), LOC(NB), LOC(EW) and LOC(EW-EWQE(v)).

Metrics. We use the Mean Reciprocal Rank (MRR) as our evaluation metric. Given a post $\mathbf{w}_i$, let the rank of its food entity be $r(\mathbf{w}_i)$, where $r(\mathbf{w}_i) = 0$ for the top rank. Over N test cases, MRR is defined as:

$$\text{MRR} = N^{-1} \sum\nolimits_{i=1}^{N} (r(\mathbf{w}_i) + 1)^{-1} \tag{9}$$

The above MRR definition is a micro measure. In a sample of test posts, more popular food entities contribute more to MRR. For further analysis, we consider treating all entities as equally important, regardless of their popularities. Thus we introduce Macro-MRR, the macro-averaged version of MRR. For all test posts related to the same food entity, we compute the MRR of the food entity. We then average the MRRs over distinct food entities. Formally:

$$\text{Macro-MRR} = E^{-1} \sum\nolimits_{i=1}^{E} \text{MRR}(e_i) \tag{10}$$

where MRR(e_i) is MRR values averaged over all test posts about entity e_i and E is the number of distinct food entities.

4.2 Results

Table 2 displays the MRR and Macro-MRR values averaged over 10 runs for each dataset. In subsequent discussions, a model is said to perform better or worse than another model only when the differences are statistically significant at p-level of 0.05 based on the Wilcoxon signed rank test.

EW and QE(v) easily outperform NB, which affirms the utility of entity-indicative weighting and venue-based query expansion. EW also outperforms QE(v), e.g. EW's MRR is 0.461 on Instagram posts, higher than QE(v)'s MRR of 0.403. By combining both models together in EWQE(v), we achieve even better performance than applying EW or QE(v) alone. This supports EWQE(v)'s modeling assumption that a word is important if it is both entity-indicative and highly related to the test post.

While venue-based query expansion is useful, user-based query expansion is less promising. Over the different datasets and metrics, QE(u) is inferior or at best on par with NB. This may be due to the entity-focused characteristic being weaker in users. This observation is consistent with our earlier empirical

Table 2. MRR and Macro-MRR values averaged over 10 runs for each dataset. The best performing model is bolded.

	Instagram		Burpple	
Model	MRR	Macro-MRR	MRR	Macro-MRR
NB	0.344	0.218	0.335	0.259
EW	0.461	0.301	0.467	0.377
QE(v)	0.403	0.236	0.389	0.252
QE(u)	0.326	0.215	0.336	0.237
EWQE(v)	0.543	0.323	0.503	0.388
EWQE(u)	0.449	0.284	0.419	0.329
NB-EWQE(v)	0.543	0.323	0.500	0.389
EW-EWQE(v)	**0.593**	**0.340**	**0.537**	**0.401**
TAGME(NB)	0.368	0.233	0.344	0.259
TAGME(EW)	0.462	0.293	0.446	0.363
TAGME(EW-EWQE(v))	0.520	0.296	0.507	0.390
LOC(NB)	0.409	0.236	0.357	0.259
LOC(EW)	0.472	0.254	0.413	0.315
LOC(EW-EWQE(v))	0.520	0.271	0.467	0.333
PTE	0.288	0.216	0.291	0.274

findings that users are less focused on food entities when compared to venues. Consequently user-based query expansion may augment test posts with noisy words less related to their food entities. Combining user-based query expansion with entity-indicative weighting also leads to mixed results. Although EWQE(u) outperforms QE(u), the former still underperforms EW.

Our results also show that the venue-based prior distribution over entities is useful, but only if it is computed from a reasonably accurate linking model. Over all dataset-metric combination, the best performing model is EW-EWQE(v) which incorporates a prior computed using the EW model. Although NB-EWQE(v) incorporates a prior as well, it utilizes the less accurate NB model. For Instagram, the tuning procedure consistently indicates in each run that the optimal η is 0 for NB-EWQE(v), thus it is equivalent to the model EWQE(v). For Burpple, the optimal η is non-zero for some runs, but NB-EWQE(v) performs only on par with EWQE in terms of statistical significance.

The TAGME variants exploit the entity-focused characteristic of venues via a voting mechanism. Performance depends on the voting mechanism as well as the underlying entity linking models. Intuitively better underlying models should lead to higher ranking accuracies in the corresponding variants. For example, TAGME(EW-EWQE(v)) outperforms TAGME(EW) while TAGME(EW) outperforms TAGME(NB). However comparing the variants against their underlying models, we note that only TAGME(NB) consistently improves over NB, while TAGME(EW) and TAGME(EW-EWQE(v)) fails to outperform EW and EW-EWQE(v) respectively. The same observation applies to the LOC variants. LOC(NB) consistently outperforms NB. LOC(EW) only outperforms EW for MRR on Instagram and is inferior in other dataset-metric combination. LOC(EW-EWQE(v)) is also inferior to EW-EWQE(v). Such mixed results of

LOC variants may be due to grid cells being less entity-focused than venues. Lastly, PTE performs poorly. We note that each entity has only one Wikipedia description page and appears in a limited number of Wikipedia contexts. Hence the Wikipedia content of food entities may be overly sparse for learning good entity representations. There are also language differences between Wikipedia pages and social media posts. This may impact cross-linking if embeddings are trained on only one source, but not the other. In conclusion, our proposed model EW-EWQE(v) performs well, despite its conceptually simple design.

4.3 Case Studies

Tables 3 to 5 illustrate different model aspects by comparing model pairs on Instagram posts. Comparison is based on the ranked position of the ground truth food entity (under column e) for each post. The ranked position is denoted as r_X for model X and is 0 for the top ranked. The ground truth entities can be inspected by appending the entity name to the URL 'https://en.wikipedia.org/wiki/'.

Table 3. Sample test posts to illustrate entity-indicative weighting. Words in larger fonts indicate larger weights under the EW model.

		e	r_{NB}	r_{EW}
S1	"#singapore we already ate claws ."	Chilli_crab	2	0
S2	"finally got to eat rojak !!!"	Rojak	5	0
S3	"#singapore #tourist "	Hainanese_chicken_rice	18	2

Entity-indicative Weighting. Table 3 compares the models NB and EW. For each test post, words with larger weights under the EW model are in larger fonts. For S1 with food entity 'Chilli_crab'[7], the largest weighted word is 'claws', referring to a crab body part. This word is rarely mentioned with other food entities, but appears in the context around the 'Chilli_crab' anchor in the Wikipedia page for 'The_Amazing_Race_25', hence it is highly indicative of 'Chilli_crab'. By assigning 'claws' a larger weight, EW improves the entity ranking over NB, from a position of 2 to 0. For S2, the word 'rojak' is indicative of the food entity 'Rojak' [8]. While NB does well with a ranked position of 5, EW further improves the ranked position to 0 by weighting 'rojak' more relative to other words. For S3, the food entity 'Hainanese_chicken_rice'[9] is described in the Wikipedia page 'Singaporean_cuisine' as the most popular dish for tourists in the meat category. Thus by assigning a larger weight to 'tourist', EW improves the linking of S3.

Query Expansion. Table 4 illustrates posts where the QE model improves over the NB model. While S4 mentions dinner, the food entity is not evident. However the word 'dinner' co-occurs with more informative words such

[7] crabs stir-fried in chilli-based sauce

[8] a traditional fruit and vegetable salad dish

[9] roasted or steamed chicken with rice cooked in chicken stock

Table 4. Sample test posts with added words (in brackets) from query expansion (QE model). The top 5 added words with largest weights are listed.

		e	r_{NB}	r_{QE}
S4	“last night dinner at #singapore #foodporn” (rice,0.25),(chicken,0.23),(late,0.21),(food,0.21),(to,0.20)	Hainanese_ chicken_rice	19	3
S5	“indian feast #daal #palakpaneer #mangolassi @rebekkariis du vil elske det!” (pakistani,0.17),(cuisine,0.17)(buffet,0.17)(lunch,0.17)(team,0.17)	Naan	1	0

Table 5. Sample test posts for comparing models EWQE(v) and EW-EWQE(v). $r_{p(e|v)}$ corresponds to ranking with the venue prior $p(e|v)$.

		e	$r_{p(e\|v)}$	$r_{EWQE(v)}$	$r_{EW-EWQE(v)}$
S6	“life’s simple pleasures. #gastronomy”	Mee_pok	0	56	0
S7	“the black pepper sauce is robust and quite spicy, one of my favourite in singapore.”	Black_ pepper_crab	1	9	2

as ‘chicken’ and ‘rice’ in other posts from the same venue. Such words are retrieved with query expansion and used to augment the post. The augmented post is then linked more accurately by the QE model. For S5, query expansion augments the post with 6 words of which 5 words share similar weights. Out of the 5 words, the word ‘pakistani’ is indicative of the food entity ‘Naan’, helping to improve the ranked position further from 1 to 0.

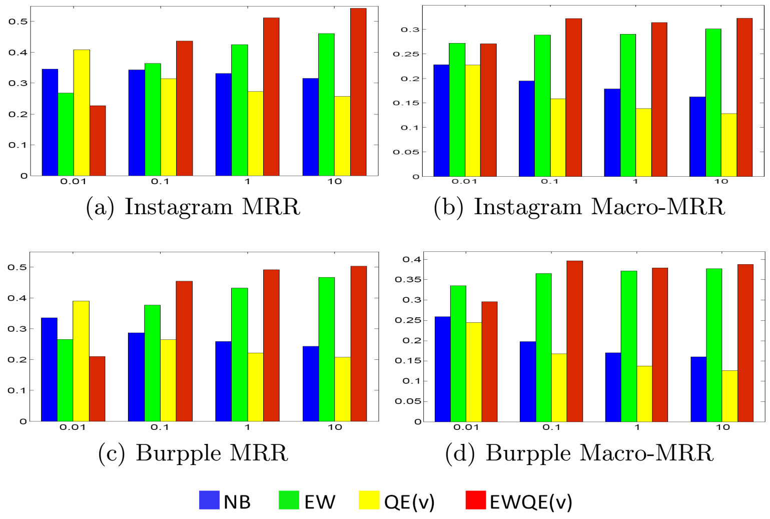
Venue-based Prior. Table 5 compares EWQE(v) and EW-EWQE(v). S6 is posted from a food venue which serves ‘Mee_pok’ [10] as one of its food entities. This food entity is mentioned explicitly in other same-venue posts. Hence on applying the EW model, we infer this venue as having a high prior probability for this entity. In fact if we rank food entities by the venue prior $p(e|v)$ alone, ‘Mee_pok’ is ranked at position 0. Integrating the prior distribution with other information as done in EW-EWQE(v), the same rank position of 0 is obtained. For S7, the ingredient black pepper sauce is mentioned, which is indicative to some extent of ‘Black_pepper_crab’ [11]. However EWQE(v) manages only a ranked position of 9. From other same-venue posts, the venue prior is computed and indicates the food entity to be highly probable at S7’s venue. Integrating this information, EW-EWQE(v) improves the ranked position to 2.

4.4 Parameter Sensitivity

For models with γ as the sole tuning parameter, we compare their sensitivity with respect to γ. Figure 2 plots the performance of NB, EW, EWQE and EWQE(v), averaged over 10 runs for different γ values. It can be seen that EWQE(v) outperforms NB over most of the applied γ values, i.e. 0.1, 1 and 10. Although EWQE(v) is simply a product combination of the EW and QE(v) models, it easily outperforms its constituent models, validating our combination approach.

[10] a Chinese noodle dish

[11] crabs stir-fried in black pepper sauce



(a) Instagram MRR

(b) Instagram Macro-MRR

(c) Burpple MRR

(d) Burpple Macro-MRR

Fig. 2. Model performance (Y-axis) with different γ values (X-axis).

This trend is consistent across both metrics and datasets. We also note that in the absence of a validation set for tuning, a natural option is to use Laplace smoothing, i.e. $\gamma = 1$. In this perfectly unsupervised setting, it is reassuring that EWQE(v) remains the best performing model. Lastly when γ is very small at 0.01, EW and EWQE(v) are under-smoothed and perform worse than NB. In this setting where smoothing is limited, QE(v) outperforms all other models, possibly because augmenting each test post with additional words is analogous to additional smoothing for selected words.

5 Related Work

Explicit Entity Linking. Compared to IEL, there has been more work in EL [10, 19, 9, 2]. In [10], Liu et al. constructed an objective function based on mention-entity features, mention-mention features etc. When linking mentions, entities are assigned via a decoding algorithm. In [2], the objective function is defined over a graph that connects tweets close in space and time. The assumption is that such tweets are likely to mention closely-related entities. In [19], Shen et al. propagate users' interest scores over an entity graph built from inter-entity semantic-relatedness [12]. Given a test mention, candidate entities with higher interest scores are preferred. Huang et al. [9] proposed label propagation over graphs with mention-entity tuples as nodes. After label propagation, high scoring tuples provide the mention-entity assignments. Finally, our baselines include extensions of EL models [5, 6]. Fang and Chang [5] learned entity distributions over time and grid cells and integrate them into a base linking system. We use their learning framework and integrate grid cell information into our model. In

[6], the idea is to let candidate entities across intra-document mentions vote for each other. For each mention, top k entities with the most votes are then filtered again by entity probability. In our extension, our voting entities are candidates for posts from the same venue, not mentions from the same document.

Implicit Entity Linking. For IEL, Perera et al. [14] built information network to link entities and knowledge nodes, using factual knowledge from the knowledge base and contextual knowledge from labeled tweets. They then use graph features to rank entities. The work in [11] engineered features from labeled tweets to train decision trees for ranking entities for each tweet. This per-tweet instead of per-mention linking resembles our IEL task. In contrast with both discussed IEL work, we assume the posts in our training set are not entity-labeled, but are associated with venues. Thus our work explores a different task setting.

Query Expansion. Query expansion originates from the document retrieval problem. To improve retrieval accuracy, potentially relevant words are weighted and added to an initial query. Cummins [4] used genetic programming to learn weighting schemes for query expansion. Qiu and Frei [16] uses a similarity thesaurus to add words that are most similar to the query concept. Xu and Croft [21] compared various query expansion techniques exploiting either the corpora-wide word occurrences/relationships or exploiting the top ranked documents returned for the initial query. Query expansion has also been applied [1] to retrieve relevant tweets given a user query. Fresno et al. [8] applied query expansion to retrieve event-related keywords. Specifically they considered candidate words from tweets close in space and time to an event-related tweet.

If we treat test posts as queries and candidate entities as documents, then IEL can be viewed as a form of document retrieval. In this work, we use query expansion to exploit the entity-focused characteristics of venues.

6 Conclusion

We have proposed novel yet well principled models for implicit food entity linking in social media posts. Our best model exploits the entity-focused characteristic of food venues and the intuition that entity-indicative words are more important for IEL, in order to outperform more complex state-of-the-art models. In future work, we intend to explore IEL in non-geotagged social media posts, where posting venues are unknown. Lastly we point out that the entity-focused characteristic appears in various forms in other problems. For example, in linking tweets to posting venues [3], users may be focused in their visits, preferring venues close to their home regions. Hence potentially, our model can be generalized to other information retrieval problems.

7 Acknowledgments

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative, and DSO National Laboratories.

References

1. A. Bandyopadhyay, M. Mitra, and P. Majumder. Query expansion for microblog retrieval. In *TREC*, 2011.
2. W.-H. Chong and E.-P. Lim. Collective entity linking in tweets over space and time. *ECIR*, 2017.
3. W.-H. Chong and E.-P. Lim. Tweet geolocation: Leveraging location, user and peer signals. *CIKM*, 2017.
4. R. Cummins. *The Evolution and Analysis of Term-Weighting Schemes in Information Retrieval.* PhD thesis, National University of Ireland, Galway, 2008.
5. Y. Fang and M.-W. Chang. Entity linking on microblogs with spatial and temporal signals. *TACL*, 2, 2014.
6. P. Ferragina and U. Scaiella. Tagme: On-the-fly annotation of short text fragments (by wikipedia entities). *CIKM*, 2010.
7. J. Ferreira, D. Denison, and D. Hand. Weighted naive bayes modelling for data mining. *Department of Mathematics, Imperial College*, 2001.
8. V. Fresno, A. Zubiaga, H. Ji, and R. Martínez-Unanue. Exploiting geolocation, user and temporal information for natural hazards monitoring in twitter. *Procesamiento del Lenguaje Natural*, 54, 2015.
9. H. Huang, Y. Cao, X. Huang, H. Ji, and C.-Y. Lin. Collective tweet wikification based on semi-supervised graph regularization. *ACL*, 2014.
10. X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu. Entity linking for tweets. *ACL*, 2013.
11. E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. *WSDM*, 2012.
12. D. Milne and I. Witten. An effective, low-cost measure of semantic relatedness obtained from wikipedia links. *AAAI*, 2008.
13. Z.-Y. Ming and T.-S. Chua. Resolving local cuisines for tourists with multi-source social media contents. *Multimedia Systems*, 22(4):443–453, 2016.
14. S. Perera, P. N. Mendes, A. Alex, A. P. Sheth, and K. Thirunarayan. Implicit entity linking in tweets. In *ESWC*, pages 118–132, 2016.
15. F. Piccinno and P. Ferragina. From tagme to wat: a new entity annotator. *ERD*, 2014.
16. Y. Qiu and H.-P. Frei. Concept based query expansion. *SIGIR*, 1993.
17. W. Shen, J. Wang, P. Luo, and M. Wang. Liege: Link entities in web lists with knowledge base. *KDD*, 2012.
18. W. Shen, J. Wang, P. Luo, and M. Wang. Linden: Linking named entities with knowledge base via semantic knowledge. *WWW*, 2012.
19. W. Shen, J. Wang, P. Luo, and M. Wang. Linking named entities in tweets with knowledge base via user interest modeling. *KDD*, 2013.
20. J. Tang, M. Qu, and Q. Mei. Pte: Predictive text embedding through large-scale heterogeneous text networks. *KDD*, 2015.
21. J. Xu and W. B. Croft. Query expansion using local and global document analysis. *SIGIR*, 1996.
22. N. A. Zaidi, J. Cerquides, M. J. Carman, and G. I. Webb. Alleviating naive bayes attribute independence assumption by attribute weighting. *JMLR*, 14(1):1947–1988, Jan 2013.