Entity Suggestion with Conceptual Explanation

Yi Zhang[1*], Yanghua Xiao[125†], Seung-won Hwang[3‡], Haixun Wang[4], X. Sean Wang[1], Wei Wang[1]
[1]Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, China
[2]Shanghai Internet Big Data Engineering Technology Research Center, China
[3]Department of Computer Science, Yonsei University, Korea
[4]Facebook Inc., USA
[5]Shuyan Technology, China
{z_yi11, shawyh, xywangcs, weiwang1}@fudan.edu.cn, seungwonh@yonsei.ac.kr, haixun@gmail.com

Abstract

Entity Suggestion with Conceptual Explanation (ESC) refers to a type of entity acquisition query in which a user provides a set of example entities as the query and obtains in return not only some related entities but also concepts which can best explain the query and the result. ESC is useful in many applications such as related-entity recommendation and query expansion. Many example based entity suggestion solutions are available in existing literatures. However, they are generally not aware of the concepts of query entities thus cannot be used for conceptual explanation. In this paper, we propose two probabilistic entity suggestion models and their computation solutions. Our models and solutions fully take advantage of the large scale taxonomies which consist of isA relations between entities and concepts. With our models and solutions, we can not only find the best entities to suggest but also derive the best concepts to explain the suggestion. Extensive evaluations on real data sets justify the accuracy of our models and the efficiency of our solutions.

1 Introduction

Entity Suggestion (*ES*) has been widely investigated in different scenarios. In a typical scenario, a system accepts a set of example entities provided by a user as a query q, and retrieves a set of entities such that these entities, along with q, complete some concepts. For example, in many online stores such as 'amazon.com', a user may browse some products such as {iPhone 6 Plus, Samsung Galaxy 6s}. It is quite possible that the user wants to buy a fashionable smart phone, then the website should recommend some other product entities such as iPhone 6s or Microsoft Lumia 950XL. Another example is when a user types {China, India, Brazil} as a query in the search engine, they may bear the concept BRIC in mind but can not recall all of its members. Thus, he/she enters these example entities of the concept for the purpose of retrieving the remaining ones. The remaining entity of BRIC, Russia should be returned as the result. We give more such examples in Table 1. *ES* is also known as *entity list completion* [Dalvi *et al.*, 2011], *entity retrieval* [Delbru *et al.*, 2012; Meij *et al.*, 2014], *entity recommendation* [Yu, 2014] or *entity query by example* [Balog *et al.*, 2011; Bron *et al.*, 2013] in different settings. *ES* has been widely and successfully used in search engine [Balog *et al.*, 2010; Mottin *et al.*, 2013], spreadsheet population, and question answering [Ahn *et al.*, 2005].

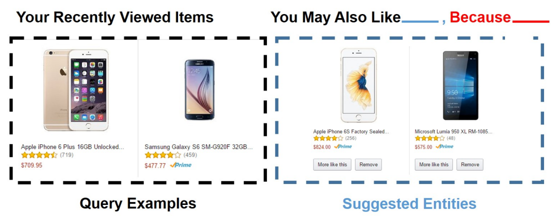

Figure 1: A typical scenario in an online store

However, *ES* can only return the suggested entities without explaining what is the meaning of the examples or why the result entities are suggested. We argue that an explanation is necessary in *ES* due to the following reasons. First, *suggested entities with reasonable explanation are more trustworthy thus encouraging more click-throughs*. In Figure 1, if we can accurately fill in the blank in red, the user can save much time to browse other suggested entities but directly browse the products he/she is interested in, and thus the website can increase the click-throughs. Second, *by providing both suggested entities and its corresponding concepts, users get more accurate feedback on whether his/her search intent was correctly recognized.* Hence, we aim to not only return semantically related entities but also provide why we suggest the entities. In this paper, we focus on the explanation by concepts of the query examples, since the concept is the major intent of users by specifying a bag of examples. We call our problem as *Entity Suggestion with Conceptual Explanation* (*ESC*).

To the best of our knowledge, this is the first work to define and provide a good conceptual explanation for entity

[*]Zhang is now affiliated with Department of Computer and Information Science, University of Pennsylvania.

[†]Corresponding author. Xiao's work was supported by National Key Basic Research Program of China (No.2015CB358800), the National NSFC (No.61472085, U1509213), Shanghai Municipal Science and Technology Commission foundation key project (No.15JC1400900), Shanghai Municipal Science and Technology project (No. 16511102102, No.16JC1420401) and Xiaoi Research.

[‡]Hwang's work was supported by IITP grant funded by the Korea government (MSIP) (No.2014-0-00147, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)).

Table 1: Examples, Suggested Entities and Explanation

Example Entities	Suggested Entities	A Possible Conceptual Explanation
China, India, Brazil	Russia	BRIC
Alibaba, Tecent	Baidu	BAT(Big Three Chinese Internet giants)
swimming, marathon	bicycle ride	Ironman Triathlon
Islam, Buddhism	Christianity	The three major religions
Standard Poor's, Moody's	Fitch Group	Big Three(credit rating agency)
Roger Federer, Rafael Nadal	Andy Murray, Novak Djokovic	Big Four(tennis)

suggestion. Many solutions of *ES* have been developed, but they can not provide such explanation. These solutions can be classified into the following three categories. The first category tends to use co-occurrence as the basic recommendation mechanism. A well-known example is Google Set whose basic idea is to recommend the entities that most frequently cooccur with the example entities. The second category assumes that the query set belongs to some list and estimate the likelihood that each item belongs to the list. An example in this category is SEISA [He and Xin, 2011]. The third category ranks all the entities based on how much their properties overlap with those of the example entities, and return the top ranked entities outside of the query as the final result [Metzger *et al.*, 2013]. However, they are generally not aware of the concepts of query entities thus are unable to conduct concept-aware suggestion, let alone give a good conceptual explanation. In our work, we think a good conceptual explanation should be both related and granularity-aware.

Recently, many web-scale conceptual taxonomies consisting of isA relationships between entities and concepts, such as Microsoft's Probase and Google's isA database, have become available. These knowledge bases are extracted by Hearst patterns from web corpora. The rich concept information in these knowledge bases brings us new opportunities to process *ESC* queries. In this paper, we use these conceptual taxonomies to find the most related entity with conceptual explanation. We propose a series of probabilistic models and approaches for concept inference and entity suggestion based on these taxonomies. We use Probase as the taxonomy, although other taxonomies can be used as well.

2 Related Work

Entity Recommendation Related entity recommendation can be categorized into the following two categories: First, to*recommend related entities for search assistance*, Blanco et al. [Blanco *et al.*, 2013] proposed a recommendation engine Spark to link a user's query word to an entity within a knowledge base and recommend a ranked list of the related entities. To guide user exploration of recommended entities, they also proposed a series of features to characterize the relatedness between the query entity and the related entities. Steffen et al. [Metzger *et al.*, 2014] proposed a similar entity search considering diversity.

Second, for *query assistance for knowledge graphs*, GQBE [Jayaram *et al.*, 2014] and Exemplar Queries [Mottin *et al.*, 2014] studied how to retrieve entities from a knowledge base by specifying example entities. For example, the input entity pair {Jerry Yang, Yahoo!} would help retrieve answer pairs such as {Sergey Brin, Google}. Both of them projected the example entities onto the RDF knowledge graph to discover result entities as well as the relationships around them. They used an edge-weighted graph as the underlying model and subgraph isomorphism as the basic matching scheme, which in general is costly. Our objective is to infer entities that preserve the semantic of the examples, thus we assume all example entities generally share the same concept.

Entity Set Expansion The goal of this line is, given a set of seed entities, to discover other entities in the same concept. Google Sets [Google, 2006] is a product implementation used to populate a spreadsheet after users provide some instances as examples. Inspired by Google Sets, many research work followed [Ghahramani and Heller, 2005; He and Xin, 2011; Wang and Cohen, 2008; Sarmento *et al.*, 2007; Pantel *et al.*, 2009], to measure the membership strength of an item for a hidden concept exemplified by query entities. However, this line of work always biases towards general concepts which is not good enough to explain the query set. Especially the query set conceptualizes to multiple fine-grained concepts, such as a camera brand and Japanese company. Our work assumes a query set can bear multiple fine-grained concepts, and aggregates a concept distribution to accurately infer related entities that reflect all the related concepts.

Related problems include semantic search tasks studied in Bron et al. [Bron *et al.*, 2013], of taking example instances and the textual representation of a relation, to complete the list of examples. Another example is harvesting tables on the Web, and retrieving the table that completes the example instances and description [He and Xin, 2011; Wang *et al.*, 2014]. Compared to this work, our task is more challenging relying solely on examples without explicit description of a relation or table.

Short Text Conceptualization Short text conceptualization aims to map a short text to a set of concepts as a mechanism of understanding text. Lee et al. [Kim *et al.*, 2013] proposed context-dependent conceptualization to capture the semantic relations between words by combining Latent Dirichlet Allocation with Probase. Song et al. [Song *et al.*, 2011] developed a Bayesian inference mechanism to conceptualize words and short texts. The ultimate objective of conceptualization is to find the concepts that best capture the semantic of the short texts. However, conceptualization combines all related concepts together without considering the semantic granularity of the concepts which increases the risk of recommending false positives and cannot explain the query well either. Our work is to find concepts that can best explain the query, and the next step of entity inference, identifying the most related entity, is beyond the scope of conceptualization.

3 Background

In this section, we briefly review the conceptual taxonomy Probase upon which our solutions are built.

Probase and isA relationships Probase [Wu *et al.*, 2011] is a universal, general-purpose, probabilistic taxonomy automatically extracted from a corpus of 1.6 billion web pages.

Probase contains 2.7 million concepts and 4.6 million isA (a.k.a., hypernym-hyponym) relationships among the concepts/entities, which is suitable for us to describe the query entities. Each isA relationship saying (e isA c) is associated with the frequency ($n(e, c)$) that e isA c is observed from the corpus. The frequency allows us to compute the *typicality* of e under concept c, i.e., $P(e|c)$, which can also be interpreted as the probability that e is an instance of c. Formally, it can be computed as follows:

$$P(e|c) = \frac{n(e, c)}{n(c)}, P(c|e) = \frac{n(e, c)}{n(e)} \quad (1)$$

where $n(c) = \sum_{e \in E} n(c, e), n(e) = \sum_{c \in C} n(e, c)$ and E, C is the whole entity and concept set in Probase respectively. We will use these two equations in the following sections.

4 Problem Model

In this section, we elaborate our two probabilistic models. We start from the simpler one.

4.1 A Probabilistic Relevance Model

Given a set of query entities $q = \{q_i | q_i \in E\}$, we model the relevance of an entity e to q with $rel(q, e)$, which can be interpreted as the likelihood that a real person will think of the entity e when he/she observes the entities in the query q. Thus, our objective is to find the entity whose relevance is the highest, and we will suggest entities based on their relevance:

$$\arg\max_{e \in E - q} rel(q, e) \quad (2)$$

Then, the key is to define the relevance function. Consider the psychological procedure of a real user to infer an entity when observing a set of example entities. The user tends to formulate the query by referring to some concepts of the examples as well as the target entity. The concepts referred to describe the one or more aspects of these entities. For example, given {China, India, Brazil}, two concepts {developing country, emerging market} naturally come to our mind. However, some other concepts such as country is also possible to be activated in our mind. But intuitively country is not as good as the other two concepts since it is more general. Hence, the key is quantifying the *promisingness* of a concept to explain and make a good entity suggestion. We use $r(c|q)$ to rank each candidate concept. Thus, we have the following relevance function:

$$rel(q, e) = \sum_i P(e|c_i) r(c_i|q) \quad (3)$$

Clearly, the relevance function is positively correlated to the two factors $P(e|c_i)$ and $r(c_i|q)$, which reflects the following two principles: (1) A *typical* entity should be suggested; (2) An entity of a *promising* concept should be suggested.

4.2 A Relative Entropy Model

An alternative model is to use a concept distribution to represent the semantics of the query entities. Thus, the entity whose admission into q can preserve the original concept distribution is exactly the entity we are looking for. More formally, let $P(C|q)$ be the concept distribution of query entity set q. The distribution can be represented as a set of vectors $\{< c_i, r(c_i|q) >\}$ (Here $r(c_i|q)$ is normalized, and can be used as a probability). We just need to find the entity e such that $P(C|q, e)$ is closest to $P(C|q)$. A popular measure of the distance between two probability distributions is *KL-divergence*, also known as *relative entropy*. Thus, our problem can be formulated as:

$$\arg\min_{e \in E - q} KL(P(C|q), P(C|q, e)) \quad (4)$$

where KL-divergence is defined as:

$$KL(P(C|q), P(C|q, e)) = \sum_{i=1}^{n} r(c_i|q) \times log(\frac{r(c_i|q)}{r(c_i|q, e)}) \quad (5)$$

4.3 Concept Ranking

The problem left is the definition of $r(c|q)$. There are two purposes to define $r(c|q)$. First, we use $r(c|q)$ (in Eq. 3) to characterize the promisingness of a concept to make a good entity suggestion. Second, we use $r(c|q)$ to select the best concepts to explain the suggestion. In other words, we use the same ranking function for two different purposes. It is reasonable, because in most cases it is the concept that help us find the right entity can best explain the suggestion. We define $r(c|q)$ as follows:

$$r(c|q) = P(c|q)\delta(c|q) \quad (6)$$

We elaborate its rationality from the following two aspects, and elaborate the computation of $P(c|q)$ and $\delta(c|q)$ in Section 5.

- **Typicality** A good concept should be the concept that people are likely to associate with the query examples. In our running example, BRIC is a very typical concept given {China, India, Brazil}, which can directly help us find the appropriate entity Russia. Thus, we define $P(c|q)$ to be the *typicality* that c is referred to when we are presented with q.
- **Granularity** A good concept should be neither too general nor too specific to summarize the query examples. In our running example, though country is a very typical concept, it is too general to explain the query entities, and it is very likely to introduce many less related entities such as Italy. Certainly, the concept can not be too specific either. To see this, the concept exactly containing the query exemplars is unable to suggest any other entities. We introduce $\delta(c|q)$ to describe the goodness of concept c in terms of its granularity given q .

5 Concept Inference

In this section, we elaborate how to compute $P(c|q)$ and $\delta(c|q)$. Finally, we discuss how to select the concepts for explanation from the ranked concept list.

5.1 $P(c|q)$ Computation

The major concern to compute $P(c|q)$ is how to aggregate the concept inference from different query entities. We propose a Naïve Bayes model and a Noisy-Or model for the aggregation. We start from a Naïve Bayes model.

Naïve Bayes Model

According to the Bayes theorem, we have:

$$P(c|q) = \frac{P(q|c)P(c)}{P(q)} \propto P(q|c)P(c) \quad (7)$$

Since $P(q)$ is only dependent on the query, it can be ignored for the purpose of ranking.

In general, a person's choices of two entities e_i, e_j are logically independent with each other given a concept c. Thus, we can have a conditional independence assumption:

$$\forall e_j, e_k \in q, P(e_j, e_k|c) = P(e_j|c)P(e_k|c) \quad (8)$$

Thus, we have:

$$\log P(c|q) \propto \log[\prod_{e_j \in q} P(e_j|c)P(c)] \propto \log(P(c)) + \sum_{e_j \in q} \log(P(e_j|c)) \quad (9)$$

Generally, there are relatively few concepts related to all of the query entities, therefore appropriate smoothing is necessary to avoid zero probabilities. To do this, we assume that with probability $1-\lambda$, the user would choose the entity by its prior typicality. Thus, Eq. 9 can be rewritten as:

$$\log P(c|q) \propto \log(P(c)) + \sum_{e_j \in q} \log(\lambda P(e_j|c) + (1-\lambda)P(e_j)) \quad (10)$$

Here the prior typicalities of $P(c)$ and $P(e_j)$ are computed by the following equations:

$$P(c) \propto n(c); P(e_j) \propto n(e_j) \quad (11)$$

where $n(c)$ (or $n(e)$) is the number of occurrence of c (or e) in Probase.

Noisy-Or Model

Alternatively, we can mimic a psychological process of identifying the concept, when query instances are presented one by one to a human—As more query entities are given, desirable concepts will amplify and eventually peak.This observation implies that the signal of the right concepts should be amplified when more entities are observed, and the signal of incorrect concepts should be weakened. These observations motivate us to use a Noisy-Or model to compute $P(c|q)$:

$$P(c|q) = 1 - \prod_{e_j \in q} (1 - P(c|e_j)) \quad (12)$$

5.2 $\delta(c|q)$ Computation

$\delta(c|q)$ evaluates the *goodness* of a concept in terms of its granularity. The key to define $\delta(c|q)$ is to punish a vague concept. We found two typical kinds of vague concepts: (1) *concepts that have many instances* and (2) *concepts that have a long distance to query entities in a conceptual taxonomy.*

Capacity-based Score

A concept with many entities might be too *general*. For example, in Probase, country has 2648 entities, while developing country has only 149 entities. Country is obviously more general than developing country. These general concepts may have high $P(c|q)$ due to their large capacity, but are relatively vague to characterize the semantics of the query entities. Hence, we define $\delta(c|q)$ as follows:

$$\delta(c|q) = \log(\frac{|E|}{Capacity(c) + 1}) \quad (13)$$

where $Capacity(c)$ is the number of entities of c, and $|E|$ is the total number of entities in Probase.

Distance-based Score

The above penalty is independent on the query q and it might bias to specific concepts. Next, we propose a new distance-based score which takes advantage of the hierarchical structure of the taxonomy and measure the penalty by the distance from query entities to the concept in the taxonomy. We use the *expected hitting time* of random walk to measure the distance. In general, an abstract concept takes a longer steps to be visited by the query entities through a random walk on the taxonomy than a specific concept does. For example, China is a country and developing country, and developing country is also a country in Probase. Thus, there exist a 2-step and an 1-step path from China to country while only an 1-step path from China to developing country. Thus, the expected hitting time of a general concept is larger. We use the inverse of expected hitting time as the penalty.

More formally, let $H(c|q)$ be the *expected hitting time* in a random walk to reach concept c starting from any entity in q along the isA relations in a taxonomy. It is the sum of $h(c|e_i)$ over $e_i \in q$. That is:

$$H(c|q) = \sum_{e_i \in q} h(c|e_i) \quad (14)$$

where $h(c|e_i)$ is the expected number of steps in a random walk starting from $e_i \in q$ to reach the concept c. In a random walk procedure, $h(c|e_i)$ is computed by the following equation:

$$\begin{cases} h(c|e_i) = 0, & \text{if } e_i = c \\ h(c|e_i) = 1 + \sum_{c' \in c(e_i)} P(c'|e_i)h(c|c'), & \text{if } e_i \neq c \end{cases} \quad (15)$$

where $c(e_i)$ is the concepts of e_i in Probase and we use $P(c'|e_i)$ (i.e., the typicality of concept c' given entity e_i) as the transition probability in the random walk procedure. Finally, we have

$$\delta(c_i|q) = \frac{1}{H(c_i|q)} \quad (16)$$

As we are only interested in the concepts within a short distance, we just ignore concepts with distance larger than a certain threshold of T steps, which reduces the computation cost. In our experiment, we set T to be 3.

A Combined Score

The above two scores use different signals for the computation. Capacity-based score might bias to specific concepts, which leads to a high precision but a low recall. Hence, a more reasonable choice is combining two scores above so that (1) all available features are used and (2) we can trade-off between precision and recall. We employ a F-score based framework for the combination. The combined score is defined as follows:

$$\delta(c|q) = (1+\beta^2)\frac{\delta_c(c|q)\delta_d(c|q)}{\beta^2\delta_c(c|q) + \delta_d(c|q)} \quad (17)$$

where δ_c and δ_d are capacity-based score and distance-based score respectively. We use β to tune the trade-off between two scores.

6 Evaluation

In this section, we systematically evaluate the effectiveness of our models and solutions with the comparison to the state-of-the-art approaches using the following two types of datasets.

- **Simple conceptual lists** A widely adopted conceptual list sets: SEAL [Wang and Cohen, 2007].
- **Specific conceptual lists** As the above dataset focuses on coarse concepts with many entities, e.g., common disease, queries, typically with a small set of examples, can match too many ground-truth concepts. To make the task more challenging, we extract a dataset of concepts with small cardinality from Wikipedia.

Table 2: Results on SEAL dataset

	2 instances					3 instances					4 instances				
Method	P@5	P@10	P	R	F	P@5	P@10	P	R	F	P@5	P@10	P	R	F
KNN	0.32	0.25	0.13	0.27	0.18	0.36	0.31	0.29	0.29	0.29	0.42	0.40	0.33	0.22	0.26
ER	0.33	0.24	0.23	0.14	0.17	0.45	0.34	0.25	0.21	0.23	0.35	0.32	0.25	0.24	0.24
ESBA	0.25	0.23	0.19	0.20	0.19	0.22	0.19	0.17	0.31	0.22	0.51	0.45	0.34	0.38	0.36
SEISA	0.41	0.39	0.33	0.33	0.33	0.46	0.43	0.35	0.38	0.36	0.43	0.38	0.32	**0.53**	0.40
PRMBA	0.60	0.44	0.41	0.35	0.38^*	0.70	0.54	0.46	0.42	0.44^*	**0.82**	0.65	0.53	0.45	0.49^*
PRMNO	**0.66**	**0.58**	**0.53**	0.47	$\mathbf{0.50}^*$	**0.76**	**0.62**	0.57	0.50	0.53^*	**0.82**	0.67	0.61	0.52	0.56^*
REMBA	0.62	0.55	**0.53**	**0.48**	$\mathbf{0.50}^*$	**0.76**	0.61	**0.59**	0.49	$\mathbf{0.54}^*$	**0.82**	**0.70**	**0.63**	0.52	$\mathbf{0.57}^*$
REMNO	0.58	0.51	0.47	0.47	0.47^*	0.74	0.59	0.52	**0.52**	0.52^*	**0.82**	0.67	0.60	0.52	0.56^*

The two models we proposed PRM (Probabilistic Relevance Model) and REM (Relative Entropy Model) can combine with two computational methods of $P(c|q)$: Naïve Bayes (BA) and Noisy-Or (NO). Hence, we have four possible combinations: PRMBA, PRMNO, REMBA, REMNO (We use the combined score of $\delta(c|q)$ in Eq. 17). We compare them with KNN, a naïve baseline, list completion solution proposed in [Adafre *et al.*, 2007] (ER), entity property based solution in [Bron *et al.*, 2013] (ESBA), and SEISA [He and Xin, 2011] (which is reported to be the best among the state-of-the-arts in [Wang *et al.*, 2014]). Our implementation of SEISA use entity lists in Probase instead of web lists which we could not access.

Furthermore, to evaluate the goodness of the conceptual explanation, we also implement the short text conceptualization method proposed in [Song *et al.*, 2011] denoted as STC. Our probabilistic relevance model and relative entropy model are used in entity suggestion, and the conceptual explanation is provided by BA and NO with the score of $\delta(c|q)$. In the comparison, we denoted them just as BA and NO.

6.1 Simple Conceptual Lists

Setup. We use English lists in the SEAL data set [Wang and Cohen, 2007] , and randomly choose 2-4 instances of each list as the query and evaluate the quality of the ranked result lists returned by different solutions with a varying number of examples.

Metrics.

- For entity suggestion. We evaluate different solutions by the following metrics: *mean Precision@k* (the ratio of correct entities that are among the top k in the ranked list divided by k, here we set k as $5, 10$), *mean precision* (the number of correct entities that are in the ranked list divided by the size of ranked list), *mean recall* (the ratio of correct entities that are in the ranked list, divided by the number of entities in the ground truth), and F_1-*score* (harmonic mean of precision and recall). We also use paired t-test to evaluate the statistic significance of the comaprision results.
- For conceptual explanation. We use the names of the lists of SEAL data set to be the ground truth of the conceptual explanation. Then we check if the ground truth concept exists in the top-k results, and return the precision. Here we report the results when $k = 5, 10$.

Results.

- Entity suggestion. The comparison results are presented in Table 2, where scores of F_1 marked with $*$ represent the winner under significance level 0.95. The results show that on SEAL data set, all of our solutions consistently outperform the competitors under all of the metrics. It also reveals that our solutions are robust against the number of given instances. These results suggest that conceptual taxonomies are beneficial for entity suggestion in general. The detailed comparisons reveal that PRMNO and REMNO in general perform better than PRMBA and PRMNO. Hence, Noisy-Or model is better than Naïve Bayes Model in concept inference.
- Conceptual explanation. The explanation results are presented in Table 5, we can see that our models outperform the method used by short text conceptualization. The BA model's better performance reveals that our consideration of the granularity of the concepts works here. Our NO model outperforms the BA shows that the Noisy-or model is better to infer the accurate concepts than the Naive Bayes model.

6.2 Specific Conceptual Lists

As motivated above, we evaluate with specific conceptual lists from Wikipedia articles. Most of these concepts have the name such as Big N and Great N. These article pages contain a list of entities that share the same specific concept, which usually requires a more complicated description. For example, from Big 4 (tennis) page in Wikipedia, we can get four entities {Roger Federer, Rafael Nadal, Novak Djokovic, Andy Murray}. The full description of the concept actually is the four most famous tennis players in the world nowadays. We collected 112 such lists representing specific concepts. Some example lists as well as their complicated concept descriptions are shown in Table 3.

Setup. Given the ground truth data, we construct the query set as follows: we choose the ground truth lists whose entities are more than 2 (all of our 112 ground truth lists have more than 2 entities), and randomly choose 2 entities as the query examples to construct the query set. In this way, we get 112 lists. Similarly, we construct another two query sets with three query entities and four entities, respectively. The size of these two data sets are 51 and 17, respectively. We run our solutions and competitors on these three query sets.

Metrics. For each query, the answer entities should be ranked higher than other unrelated entities. Thus, on this data set, we use *mean NDCG* to evaluate each query in q, denoted as *mNDCG*. Obviously, a larger mNDCG implies a better ranking. Since mNDCG evaluates the precision of our results, we further use *mean Recall@k* to evaluate the recall of our approaches. Here, we denoted it as *mRecall@k* and set k from 1 to 10.

Results The results measured by mNDCG are shown in the last figure of Figure 2. It is evident that our approaches are better than the baseline and other competitors. For example, our mNDCG scores are significantly higher than SEISA at the significant level higher than 0.96. The result can be consistently observed no matter how many seeds are provided. It also reveals the performance of our four methods are quite close to each other on this data set.

In first three figures of Figure 2, we show the results under the measurement of mRecall. We can see our approaches perform better than the competitors consistently over different k and different number of seeds. We highlight that for more

Table 3: Entity lists of specific concepts

Entity List	The full description of the concept
Leonnado da vinci, Raphael, Michelangelo	the three most famous art masters of the renaissance
Aaron Kwok, Jacky Cheung, Leon Lai, Andy Lau	the four most famous singers in hongkong
Cats, Miss Saigon, Les Miserables, Phantom of the Opera	the four most famous and classical musicals
Agricultural Bank of China, China Construction Bank, Bank of China, Industrial and Commercial Bank Of China	the four biggest banks owned by chinese government

Table 4: Top-5 concepts

Query 1: China, India, Brazil		**Query 2: PricewaterhouseCoopers, KPMG**		**Query 3: Aaron Kwok, Jacky Cheung**		**Query 4: Tencent, Baidu**	
NO	STC	NO	STC	NO	STC	NO	STC
emerging market	country	global business consultancy firm	company	hot Hong Kong singer	performer	Chinese internet giant	good company
developing country	economy	big accounting firm	firm	popular canto-pop entertainer	idol	Shanghai-Chinese internet giant	Chinese company
emerging market	developing country	accounting firm	global consultant	Hong Kong artist	spokespersons	Chinese technology company	big company
country	nation	large audit firm	respected auditing firm	universal's other front-line artist	popular canto-pop entertainer	Chinese social networking site	site
economy	market	well-known auditing company	financial accounting firm	famous celebrity guest	hot Hong Kong singer	mega-player	company

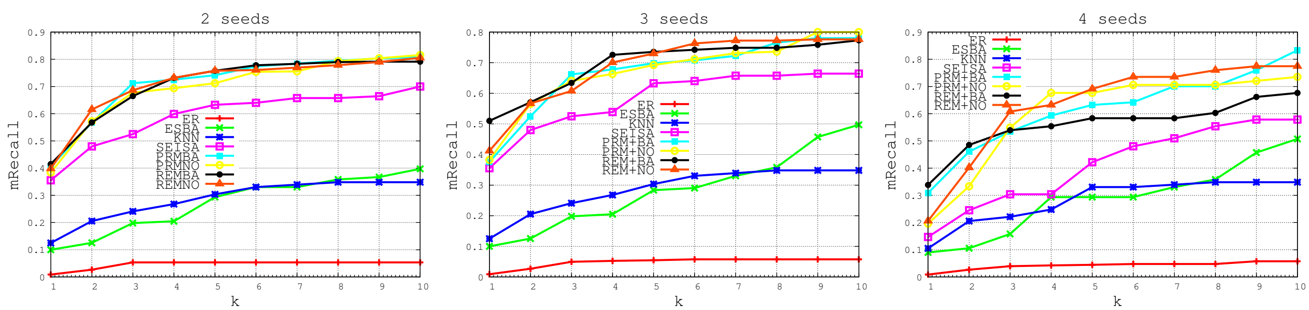

Figure 2: Mean recall@k and NDCG of different solutions

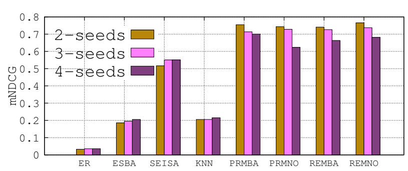

Figure 3: NDCG of different solutions

than 70% (twice of the competitors) queries, our approaches can find the right entity in the top-6 candidates.

Case Study When doing entity suggestion on the specific conceptual lists, the good conceptual explanation should be fine-grained, but it may be different with the name of the list from Wikipedia. Thus, human evaluation should be involved, and we provide the case study of the results instead of the study of the precision here.

In Table 5, we give case studies to show the effectiveness of our conceptual explanation. Because of the space limitation, Table 4 presents the top-5 concepts of 4 queries using NO and STC. Since we have shown that NO is better than BA to provide explanation, here we omit the result of BA. The query examples are in the first row of the table. According to the results, we can see that the concepts founded by STC are quite vague. However, the concepts found by NO are much more specific and related.

The above results sufficiently show that the careful selection of concepts is critical for both the effectiveness of concept selection and entity inference.

Table 5: Precision of the conceptual explanation

k	STC	BA	NO
5	0.091	0.273	**0.545**
10	0.182	0.455	**0.727**

7 Conclusions

This paper studies entity suggestion with conceptual explanation, a technique with diverse applications. Specifically, we have proposed two probabilistic approaches, the first leveraging the typicality of concepts and entities to make the inference biased toward the more promising entities, and the second using an optimization solution to minimize the difference before and after the acceptance of a candidate entity. With these two models, we have solved the challenging problems of how to aggregate the conceptual information of the example entities by a Naïve Bayes Model and a Noisy-Or model and how to find specific concepts by a capacity-based approach and a distance-based approach. We have validated the effectiveness of our approaches using extensive evaluations with real-life data.

References

[Adafre *et al.*, 2007] S Fissaha Adafre, Maarten de Rijke, and E Tjong Kim Sang. Entity retrieval. *Recent Advances in Natural Language Processing (RANLP 2007)*, 2007.

[Ahn *et al.*, 2005] Kisuh Ahn, Johan Bos, David Kor, Malvina Nissim, Bonnie L Webber, and James R Curran. Question answering with qed at trec 2005. In *TREC*, 2005.

[Balog *et al.*, 2010] Krisztian Balog, Pavel Serdyukov, and Arjen P de Vries. Overview of the trec 2010 entity track. Technical report, DTIC Document, 2010.

[Balog *et al.*, 2011] Krisztian Balog, Marc Bron, and Maarten De Rijke. Query modeling for entity search based on terms, categories, and examples. *ACM Transactions on Information Systems (TOIS)*, 29(4):22, 2011.

[Blanco *et al.*, 2013] Roi Blanco, Berkant Barla Cambazoglu, Peter Mika, and Nicolas Torzec. Entity recommendations in web search. In *The Semantic Web–ISWC 2013*, pages 33–48. Springer, 2013.

[Bron *et al.*, 2013] Marc Bron, Krisztian Balog, and Maarten De Rijke. Example based entity search in the web of data. In *Advances in Information Retrieval*, pages 392–403. Springer, 2013.

[Dalvi *et al.*, 2011] Bhavana Dalvi, Jamie Callan, and William Cohen. Entity list completion using set expansion techniques. Technical report, DTIC Document, 2011.

[Delbru *et al.*, 2012] Renaud Delbru, Stephane Campinas, and Giovanni Tummarello. Searching web data: An entity retrieval and high-performance indexing model. *Web Semantics: Science, Services and Agents on the World Wide Web*, 10:33–58, 2012.

[Ghahramani and Heller, 2005] Zoubin Ghahramani and Katherine Heller. Bayesian sets. 2005.

[Google, 2006] Google. Googlesets. URL:http://labs.google.com/sets, accessed on 04-10-2006, 2006.

[He and Xin, 2011] Yeye He and Dong Xin. Seisa: set expansion by iterative similarity aggregation. In *Proceedings of the 20th international conference on World wide web*, pages 427–436. ACM, 2011.

[Jayaram *et al.*, 2014] N. Jayaram, M. Gupta, A. Khan, Chengkai Li, Xifeng Yan, and R. Elmasri. Gqbe: Querying knowledge graphs by example entity tuples. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 1250–1253, March 2014.

[Kim *et al.*, 2013] Dongwoo Kim, Haixun Wang, and Alice Oh. Context-dependent conceptualization. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2654–2661. AAAI Press, 2013.

[Meij *et al.*, 2014] Edgar Meij, Krisztian Balog, and Daan Odijk. Entity linking and retrieval for semantic search. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 683–684. ACM, 2014.

[Metzger *et al.*, 2013] Steffen Metzger, Ralf Schenkel, and Marcin Sydow. Qbees: query by entity examples. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1829–1832. ACM, 2013.

[Metzger *et al.*, 2014] Steffen Metzger, Ralf Schenkel, and Marcin Sydow. Aspect-based similar entity search in semantic knowledge graphs with diversity-awareness and relaxation. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01*, pages 60–69. IEEE Computer Society, 2014.

[Mottin *et al.*, 2013] Davide Mottin, Themis Palpanas, and Yannis Velegrakis. Entity ranking using click-log information. *Intelligent Data Analysis*, 17(5):837–856, 2013.

[Mottin *et al.*, 2014] Davide Mottin, Matteo Lissandrini, Yannis Velegrakis, and Themis Palpanas. Exemplar queries: Give me an example of what you need. *Proceedings of the VLDB Endowment*, 7(5), 2014.

[Pantel *et al.*, 2009] Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu, and Vishnu Vyas. Web-scale distributional similarity and entity set expansion. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 938–947. Association for Computational Linguistics, 2009.

[Sarmento *et al.*, 2007] Luis Sarmento, Valentin Jijkuon, Maarten de Rijke, and Eugenio Oliveira. More like these: growing entity classes from seeds. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 959–962. ACM, 2007.

[Song *et al.*, 2011] Yangqiu Song, Haixun Wang, Zhongyuan Wang, Hongsong Li, and Weizhu Chen. Short text conceptualization using a probabilistic knowledgebase. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, IJCAI'11, pages 2330–2336. AAAI Press, 2011.

[Wang and Cohen, 2007] Richard C Wang and William W Cohen. Language-independent set expansion of named entities using the web. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 342–350. IEEE, 2007.

[Wang and Cohen, 2008] Richard C Wang and William W Cohen. Iterative set expansion of named entities using the web. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 1091–1096. IEEE, 2008.

[Wang *et al.*, 2014] Chi Wang, Kaushik Chakrabarti, Yeye He, Kris Ganjam, Zhimin Chen, and Philip A Bernstein. Expansion of tail concept using web tables. 2014.

[Wu *et al.*, 2011] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Zhu. Towards a probabilistic taxonomy of many concepts. Technical report, 2011.

[Yu, 2014] Xiao Yu. *Entity recommendation and search in heterogeneous information networks*. PhD thesis, University of Illinois at Urbana-Champaign, 2014.