

OPU at NTCIR-11 RecipeSearch - Japanese Recipe Pairing by Naïve Bayes Estimation with Names of Ingredients

Yasuhiro Tajima

Genichiro Kikui

Megumi Kubota

Rikako Inoue

Okayama Prefectural University

111, Kuboki, Soja, Okayama

7191197 Japan

+81-866-94-2128

tajima@cse.oka-pu.ac.jp

ABSTRACT

We suggest a naïve Bayes method for Japanese recipe pairing which uses ingredients of main and side dishes. For every pair of ingredients in the learning data, we calculate the probability of the co-occurrence. For a main dish of the evaluation, we guess a side dish whose posterior probability has the maximum value. In our experiment, the domain of evaluation data is restricted to the 100 examples which are given by TOs. When we evaluate a main dish, we calculate the posterior probability for every side dish in the 100 examples. Then the side dish whose posterior probability is the maximum is guessed.

Team Name

OPU

Subtasks

RecipeSearch JA 2

Keywords

naïve Bayes, ingredients list, recipe pairing

1. SYSTEM OVERVIEW

This system works by the two stages:

1. the learning phase and
2. the evaluation phase.

In the learning phase, we calculate the co-occurrence probability of a pair of two ingredients. One of the ingredients is from main dish and the other is from a side dish. In the evaluation phase, the system takes ingredients of a main dish as input. We can find the posterior probability for a side dish from the co-occurrence probability calculated by the learning phase. Then it is guessed by our system that the side dish which has the maximum posterior probability.

We evaluate the following two cases of learning data.

1. leave-one-out test

The learning data consists of 99 entries of example data. The

rest of one entry will be used as the evaluation data. Every entry of examples contains the ingredients of main dish and side dish. For example, an entry of examples is as follows:

```
main["肉類/ぶた/こまぎれ","りんご","ワイン/白","牛乳","マスタートド/粒","塩","しょうゆ"],side["じゃがいも","たまねぎ","長ねぎ","だし類/ブイヨン/チキン","豆乳","牛乳","バター","塩","こしょう","だし類/コンソメ","粉寒天","パセリ"]
```

We call one of ingredient in the entry “item”. Every item corresponds to an ingredient in [1].

2. closed test

The learning data consists of all 100 entries of example data. The evaluation data is also one entry which is selected from the learning data.

2. FINDING THE POSTERIOR PROBABILITY

Suppose that ingredients of a main dish M which has i items is

$$M(m_1, m_2, \dots, m_i)$$

In addition, suppose that the ingredients of a side dish N which has j items is

$$N(n_1, n_2, \dots, n_j)$$

Let the following be an entry of the learning data:

$$E(m_1, m_2, \dots, m_i, n_1, n_2, \dots, n_j)$$

The posterior probability of an item of the side dish n_k given m_l is

$$P(n_k | m_l) = \frac{P(m_l | n_k)P(n_k)}{P(m_l)}$$

and the posterior probability of N given M is

$$P(N | M) = \frac{\prod_{k=1, l=1}^{k \leq j, l \leq i} P(n_k | m_l)}{j}$$

In the learning phase, we must calculate

$$P(m_l | n_k)$$

for all pair of m_l, n_k ($1 \leq l \leq i, 1 \leq k \leq j$).

We use the prior probability $P(n_k)$ as a uniform probability:

$$P(n_k) = \frac{1}{j}.$$

3. EVALUATION

For evaluation, we define two types of items.

1. Using core name for items:

If the item is separated by “/” then the first part of the separation is used for the item name. For example, “マスタード/粒” is converted to “マスタード”.

But, if the first part of the separation contains “類” then the second part of the separation is used for the item name. For example, “肉類/ぶた/こまぎれ” is converted to “ぶた”.

2. Using exact item name:

All item names are used as they are. For example, “肉類/ぶた/こまぎれ” or “マスタード/粒” are used for item names.

The results of run were made by the following conditions.

OPU-JA2-ORCL-01 : (1)leave-one-out test and (1)core name for items are used.

OPU-JA2-ORCL-02 : (2)closed test and (1)core name for items are used.

OPU-JA2-ORCL-03: (1)leave-one-out test and (2)exact item name are used.

OPU-JA2-ORCL-04 : (2)closed test and (2)exact item name are used.

Each setting has marked the following precisions[2].

OPU-JA2-ORCL-01 : 0.01

OPU-JA2-ORCL-02 : 0.55

OPU-JA2-ORCL-03 : 0.00

OPU-JA2-ORCL-04 : 0.94

From these results, the closed test can find the correct items but open test setting cannot find correctly. This might be because of lack of learning data.

4. FEATURES

In this experiment, we only use names of ingredients. For the future work, the following features might be used.

- Quantity of ingredients

The main ingredients will affect the taste of dishes. Of course, quantity is important to decide the main ingredients of dishes.

- Cooking process

There are some mismatch combination of dishes, for example, fries and juicy fruit. Using the text of cooking process, we can avoid such combinations.

5. REFERENCES

[1] STANDARD TABLES OF FOOD COMPOSITION IN JAPAN - 2010 - Report of the Subdivision on Resources, The Council for Science and Technology Ministry of Education, Culture, Sports, Science and Technology, JAPAN

[2] Michiko Yasukawa, Fernando Diaz, Gregory Druck and Nobu Tsukada "Overview of the NTCIR-11 Cooking Recipe Search Task," in Proceedings of NTCIR-11, 2014.

6. ACKNOWLEDGEMENT

This work is supported by The Rakuten Data provided by Rakuten, Inc.