4F1-OS-5-4

Semi-supervised Learning for Recognizing Factors in Argumentation Support Systems

Youwei Lu[*1] Shogo Okada[*1] Katsumi Nitta[*1]

[*1]Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering Tokyo Institute of Technology

In this paper we have introduced a semi-supervised Naïve Bayes classifier to recognize factors in one`s utterance which are elementary units in argumentation support systems such as online training support systems for ADR mediator, deliberation process support systems and etc. We also show that the semi-supervised classifier making use of labeled and unlabeled data empirically outperforms the supervised in inference of factors contained in new utterances. Future efforts for extension of the model are also discussed.

1. Introduction

Argumentation is an interactive and communicative activity, and it exists extensively in our social life. Taking advantage of modern technology, we have achieved not only more convenient and economic communication but also more time-consuming argumentations, and thus better argumentation support is desired.

To equip law students with better mediation skills, an online training support system for ADR mediators has been developed in [Tanaka 2005], which serves as a powerful tool for the supervisor to instruct several students at the same time. [Sato 2011] has proposed a system to support the deliberation process for citizen judge trials.

Despite different application backgrounds, what these argumentation support systems have in common is that the logical propositions, which are called factors, involved in the case are studied and extracted in advance, and that the support given by systems is based on comparing and analyzing the combinations of factors contained in one's utterance. An illustrative list of factors, with which we have conducted experiments in Section 4, is shown in Table 1.

Obviously it is impossible to compare different argumentations if the logic propositions contained in them differ with one another. Therefore, factors are introduced here, and they are shared among different argumentations. Since each argumentation is represented in terms of the same factors, the comparison between different argumentations can be made, and thus support can be provided by argumentation support systems. For example, suggestions can be made by the system for the users to reach a compromise more efficiently. The concept of factors discussed here is very similar to that of dimension in [Ashley 1991] and factor in [Aleven 1996].

However, the recognition for factors contained in the natural language is far from automatic in existing argumentation support systems, which means one has to manually select appropriate factors for each utterance. It is quite a tedious task because the argumentation has to be interrupted for the selection of factors.

Contact : Lu Youwei , 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, royui@ntt.dis.titech.ac.jp

Exacting factors from natural language is a complex process, and many approaches and models have been proposed [Mori 2008]. Faced with complicated natural language, they are either difficult to implement or unable to perform well.

In this paper, a semi-supervised Naïve Bayes classifier is used to recognize what factors the users are talking in argumentation support systems in, and the experimental results empirically prove its capability of inference with certain accuracy.

2. Factor Recognition in Argument Support Systems

2.1 An overview of argumentation support systems

Argumentations happen with at least two people, and it is quite straightforward to extend the system involving only two to the multiuser. So we illustrate the implementation of the system with only two users in the Figure 2. As Figure 2 shows, instead of analyzing the utterances directly from users, argumentation support systems firstly extract factors from a user's utterance in the stage of factor recognition. And then represented in terms of factors, the natural language is "understood" by the system according to the combination of factors. Since factors are all pre-defined, related factors, which once have been stated in other argumentations, can be searched throughout the database in the system. Having acquired existing factor-related information in previous argumentations, logic processing for argumentation support is activated, and thus advice can be provided under the present argumentative circumstance by the system as a feedback. For example the system proposed in [Tanaka 2005] can help users come to a compromise more efficiently, and the system developed in [Sato 2011] supports the deliberation process for citizen judge trials.

2.2 Factor Recognition

Here we will show the implementation of a semi-supervised Naïve Bayes classifier and how to utilize it to predict with real data. Figure 3 gives a flow diagram of processing and the illustrative images of outcome in each processing stage.

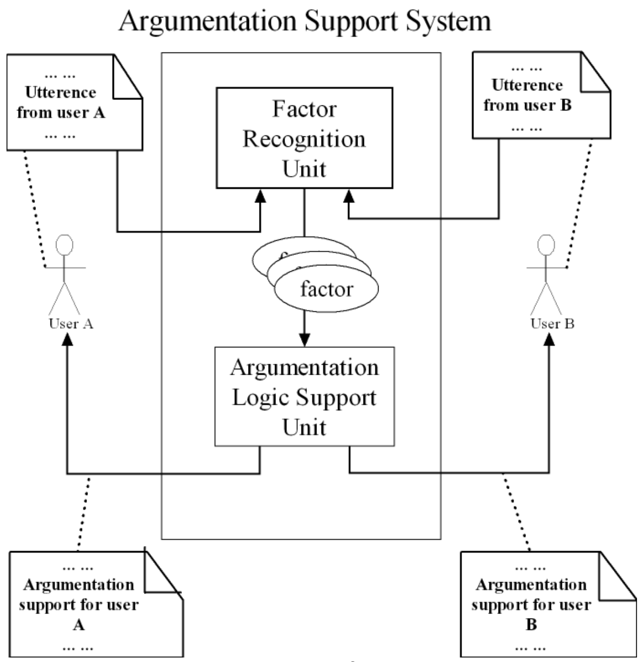

Figure 2: an illustrative graph of argumentation support systems involving two users.

Firstly, a factor list must be made. Since factors are the kernel in the factor recognition, the factor list must be made very carefully based on the existing cases before building a semi-supervised Naïve Bayes classifier. Secondly, we pick up part of utterance data randomly as learning data, and classify each utterance into the group of corresponding factor, or attach each of them the label of corresponding factor according to the factor list. Although only part of data is used as learning data which consists of the labeled data now, not only the labeled but also the unlabeled data used for learning is to be transferred to vectors using a morphological parser called ChaSen. Each element of a vector counts appearance of a particular word. Fourthly, making use of the vectors obtained from the labeled and unlabeled data, we train a semi-supervised Naïve Bayes classifier, which will be discussed in Section 3. Lastly, recognition tests are conducted. Of course morphological analysis may be needed again.

3. Semi-supervised Naïve Bayes Model

The idea of semi-supervised Naïve Bayes learning using EM algorithm is proposed in [Nigam 2006]. We use it as a power tool to infer what factors corresponds to the users' words. We introduce the algorithm briefly here and give detailed calculation of estimate of parameter using EM algorithm.

3.1 Notation

Semi-supervised Naïve Bayes Model is a generative model. It assumes the generative process for each utterance d_i as follow:

- Choose the length of an utterance, denoted by N_i with an uniform distribution.
- Choose a factor z_i for the utterance according to a multinomial distribution.
- For each w_n in the utterance out of N, choose a word according to the multinomial distribution with respect to z_i.

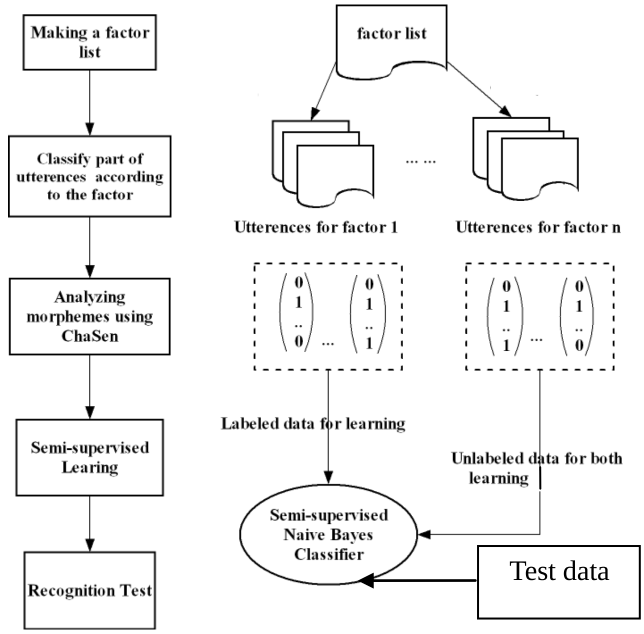

Figure 3: flow diagram for factor recognition (left), and illustrative images of outcome in each processing stage (right).

The corpus C consist of two parts: a collection of labeled utterances X_l with label vector Y_l, and another unlabeled X_u.

3.2 Supervised Learning

Training a supervised naive Bayes classifier consists of estimating the parameter of the model. Maximum likelihood estimation (MLE) yields the estimate of parameters in supervised learning as follows:

$$p(z_j) \propto \sum_i \delta(d_i, z_j) \text{ ,} \tag{1}$$

$$p(w_m \mid z_j) \propto \sum_i \delta(d_i, z_j) n(d_i, w_m) \text{ ,} \tag{2}$$

where

$$\delta(d_i, z_j) = \begin{cases} 1 & \text{when } d_i \text{ is classified as } z_j, \\ 0 & \text{otherwise.} \end{cases}$$

3.3 Semi-supervised Learning with EM algorithm

Although the calculation of supervised Naïve Bayes is simple and has an intuitive appeal, it is tedious even impossible to attach each of them a label manually when the amount of unlabeled data is huge. Thus semi-supervised learning can be used

The semi-supervised learning for naïve Bayes with EM algorithm is as follows:

- Train an initial naïve Bayes classifier from X_l by maximum likelihood estimation (MLE).
- Maximize the overall likelihood for the corpus using EM algorithm until a stable classifier converges.
  - (E step) Use the current classifier to evaluate the likelihood of unlabeled documents X_u.
  - (M step) Refine the classifier by updating the parameters which maximize the overall likelihood for the corpus.

3.4 Calculation of the semi-supervised Learning with EM algorithm

We give detailed calculative procedure of the EM algorithm here. Having divided the corpus C into the labeled and the unlabeled part, we calculate the overall likelihood as follow.

$$\begin{aligned} l &= \log P(X_l, Y_l, X_u) \\ &= \log P(X_l, Y_l) + \log P(X_u) \\ &= \sum_{d_i \in X_l} \sum_j \delta(d_i, z_j) \log p(z_j) \\ &+ \sum_{d_i \in X_l} \sum_j \sum_m \delta(d_i, z_j) n(d_i, w_m) \log p(w_m \mid z_j) \\ &+ \sum_{d_i \in X_u} \log \sum_j P(z_j) \prod_m p(w_m \mid z_j)^{n(d_i, w_m)} \end{aligned}$$

Note that the third term in the equation above has a lower bound according to Jensen's inequality.

$$\begin{aligned} &\sum_{d_i \in X_u} \log \sum_j P(z_j) \prod_m p(w_m \mid z_j)^{n(d_i, w_m)} \\ &\geq \sum_{d_i \in X_u} \sum_j Q_i(z_j) \log P(z_j) \prod_m p(w_m \mid z_j)^{n(d_i, w_m)} / Q(z_j) \\ &= \sum_{d_i \in X_u} \sum_j Q_i(z_j) \left[\log P(z_j) + \sum_m n(d_i, w_m) \log p(w_m \mid z_j) \right] + const, \end{aligned}$$

where

$$Q_i(z_j) = P(z_j \mid d_i) \propto P(z_j) \prod_m P(w_m \mid z_j)^{n(d_i, w_m)}.$$

Note the following constraints always hold:

$$\sum_j p(z_j) = 1,$$

$$\sum_m p(w_m \mid z_j) = 1.$$

Adding corresponding Lagrange multiplier, we obtain equations below to maximize the overall likelihood.

$$p(z_j) = \frac{\sum_{d_i \in X_l} \delta(d_i, z_j) + |X_u| \sum_{d_i \in X_u} Q_i(z_j)}{|X_l| + |X_u|}, \qquad (3)$$

$$p(w_m \mid z_j) = \frac{\sum_{d_i \in X_l} \delta(d_i, z_j) n(d_i, w_m) + \sum_{d_i \in X_u} Q_i(z_j) n(d_i, w_m)}{\sum_{d_i \in X_l} \sum_m \delta(d_i, z_j) n(d_i, w_m) + \sum_{d_i \in X_u} \sum_m Q_i(z_j) n(d_i, w_m)} \qquad (4)$$

4. Experimental Results

With a collection of argumentation log data from a specific case, we have extracted 17 factors out of the case, and have prepared a word list consisting of 117 words for utterance recognition. A slice of the factor list is shown in Table 1. The corpus consists of 177 utterances. We randomly select a certain number of the utterances as labeled data, the rest as unlabeled data for both learning and testing. An illustrative slice of testing records is shown in Table 2.

The empirical results are shown in Diagram 1. The X axis in Diagram 1 represents the number of labeled data, and the Y axis represents the accuracy the classifier obtain in inference of the test data.

As we can see from Diagram 1, the accuracy given by both the semi-supervised and the supervised increases as more labeled data is used, and the performance of the semi-supervised exceeds that of the supervised greatly especially when the labeled data is very limited in quantity compared to the whole corpus. For example, with 38 labeled data the semi-supervised reaches more than 60%, where only 20% data in the corpus is labeled. Besides, it can be seen that the more labeled is used, the closer the supervised gets to the semi-supervised. This is very straightforward if we compare (3)(4) with (1)(2). As the labeled data increases, its impact on the semi-supervised learning becomes dominant, and then (3)(4) will get closed to (1)(2).

Table 1: a slice of the factor list used in the experiment

F1	ステンレス製ではなくアルスター製 (It is made of aluminum rather than stainless steel.)
F2	アルスターがステンレスより安い (The aluminum is cheaper than the stainless-steel.)
F3	製造番号の刻印がない (There is no product no.)
F4	特注品である (It is a custom-made product.)
F5	商品の写真を載せてある (The photo is displayed.)
F6	パイプがアルスターであるかは明記されていない (There is no clear statement that it is made of aluminum)
F7	HP にはアルスター製がラインナップにない (No product made of aluminum is shown on the web page)
F8	現在生産されているものはステンレス製のみ (Only the stainless-steel is in production)

Table 2: a slice of recognition test records by the semi-supervised Naïve Bayes classifier using 76 labeled data

User's utterance	Extracted keywords	Factor	Inference
…, 私の方ではマフラーの説明はなかったですね. (There is no clear description about the product, isn't?)	説明 (description) ない (no)	7	7:77.06% 17:22.94%
…マフラーはステンレス製の方はずいぶんと値段が高いそうですが, Y さんは…. (The stainless-steel is much expensive than the aluminum.)	買う (buy) ステンレス (stainless) 高い (expensive)	2	1:71.14% 8:28.86%
…,Y さんがメールを送ったのは発送から 2 ヶ月半経ってからということだったんですが,… (I didn`t receive your letter of complaint until two months have passed by.)	ない (no) 経つ (pass) 二ヶ月 (two months)	16	16:88.48% 17:11.52%

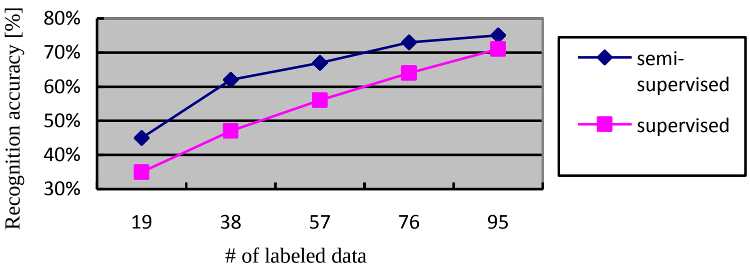

Figure 4: accuracy of inference by semi-supervised and supervised learning

5. Discussions and Future Works

In this paper, we introduce a semi-supervised Naïve Bayes classifier to recognize factors in argumentation support systems. The experimental results reveal that it works better than the supervised in inferring corresponding factors from one's utterance.

However, the Naïve Bayes model assumes words are generated after the type of the document is decided. So when several factors exist in the utterance at the same time, sentence segmentation is still needed manually in our system. How to infer the factors without sentence segmentation remains the future works.

References

[Tanaka 2005] Takahiro Tanaka, Yosiaki Yasumura, Daisuke Katagami and Katsumi Nitta: Case Based Online Training Support System for ADR Mediator, Proc. of International Conference on Artificial Intelligence and Law (ICAIL), 2005.

[Sato 2011] Takanori Sato, Syogo Okada and Katsumi Nitta: Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors, Proc. 5th International Workshop on Juris-informatics (JURISIN 2011), 2011.

[Ashley 1991] Kevin D. Ashley: Reasoning with cases and Hypotheticals in HYPO, In: Int. J. ManMachine Studies, 1991.

[Aleven 1996] Vincent Aleven , Kevin D. Ashley: How Different Is Different? Arguing about the Significance of Similarities and Differences, In Advances in Case-Based Reasoning: Proceedings EWCBR-96, 1996.

[Mori 2008] De Mori and R. Bechet F: Spoken language understanding. Signal processing magazing, IEEE, 2008.

[Nigam 2006] Kamal Nigam, Andrew Mccallum, Tom Mitchell: Semi-supervised Text Classification Using EM, Semi-Supervised Learning, 2006.