



# Bayesian Diagnosis Tracing: Application of Procedural Misconceptions in Knowledge Tracing

Junchen Feng<sup>(✉)</sup>, Bo Zhang, Yuchen Li, and Qiushi Xu

17zuoye, Chaoyang District, Beijing 100020, China  
junchen.feng@17zuoye.com

**Abstract.** Bayesian diagnosis tracing model (BDT) replaces the generic “wrong” response in the classical Bayesian knowledge tracing model (BKT) with a vector of procedure misconceptions. Using a novel dataset with actual student responses, this paper shows the BDT model has better interpretability of the latent factor and minor improvement in out-sample predictability in some specification than the BKT model.

**Keywords:** Procedural misconception · Bayesian knowledge tracing · Bayesian diagnosis tracing · Hidden Markov model

## 1 Introduction

### 1.1 Motivation

In our frequent exchanges with front-line teachers, a question often arises: “What does the 84% mastery mean in reality? Could you show us what students actually submitted?” Teachers are not only interested in predicting whether a student gets a question wrong, but also how they get it wrong. For example, the most frequent wrong answer to  $54 - 26$  is 38: students forget to trade a ten from the digit in tens. A less frequent wrong response is 32, which is caused by misunderstanding the rule of decomposition and treat the larger number in each digit as minuend ( $5 - 2 = 3$ ,  $6 - 4 = 2$ ). The latter error exposes a more critical procedure misconception of subtraction. However, The Bayesian Knowledge Tracing (BKT) model (Corbett and Anderson [1]) cannot answer the question of “how” because of an implicit assumption that the response is a binary variable, thus all wrong responses are qualitatively the same.

### 1.2 Literature Review

Pelánek and Desmarais both provid the latest literature review on this extending the BKT model [2, 3]. Among them, the most influential innovations are contextual slip and guess parameter (Baker et al. [4]), individualized model (Yudelson et al. [5], Pardos and Neil [6]), and Deep knowledge tracing (Piech et al. [7]). Instead of elaborating the latent factor structure, this paper proposes to enlarge the observations. Such idea draws inspirations from VanLehn [8]’s work on procedure misconceptions. Liu et al. [9]

encodes the misconception in the structures of knowledge components. In contrast, this paper treats the misconceptions as observable responses.

## 2 Diagnosis of Procedure Misconceptions

### 2.1 Dataset

The dataset comes from the Optical Character Recognition (OCR) of mental arithmetic practice booklet. Mental arithmetic means no vertical procedure. A student writes the answer on the booklet and takes a photo. An app auto-mark the photographed booklet so that a teacher does not need to. Figure 1 is a screenshot of a marked booklet.

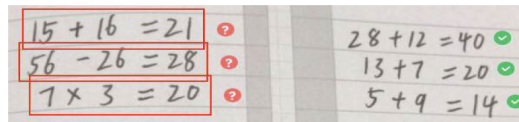


Fig. 1. The OCR of an mental arithmetic practice booklet

The paper extracts two-digit subtraction items from the OCR data submitted during December 2018. It excludes students who practiced less than 5 times or more than 200 times. The remaining dataset includes 627,330 practices from 22,395 students, with a correct percentage of 92%.

### 2.2 Misconception Diagnosis

This paper identifies the following procedure misconceptions: forget borrowing a ten ( $54 - 26 = 38$ ), miss one ( $54 - 36 = 27/39$ ), miss the digit of tens ( $54 - 36 = 8$ ) and general misconception of subtraction. The last category includes unnecessary trading a ten from the next digit ( $56 - 24 = 22$ ) and treating larger number as the minuend in each digit ( $54 - 26 = 32$ ). “skip” is not procedure misconceptions but frequent enough to merit its own category: leave a line empty (“ $54 - 26 = \_$ ”) or fill it with a number from the expression (“ $54 - 26 = 26$ ”). Table 1 lists the distribution of wrong responses.

Table 1. Distribution of wrong responses

Pattern	Example	Percentage
Skip	$54 - 26 = \_$ ; $54 - 26 = 54$	12.4%
Forget Borrowing a Ten	$54 - 26 = 38$	13.5%
Miss One	$54 - 26 = 27$ ; $54 - 26 = 29$	8.9%
General Misconception	$54 - 26 = 32$ ; $56 - 24 = 22$	7.1%
Miss the Digit of Tens	$54 - 26 = 8$	3.2%
No-diagnosis	$54 - 26 = 1$ ;	54.8%

It should be noticed that more than half of the wrong responses are not diagnosed: Even for such a quite simple arithmetic operation, the distribution of misconceptions has a very long tail.

### 3 Bayesian Diagnosis Tracing Model

The misconception-as-observation model is called Bayesian Diagnosis Tracing Model (BDT), to distinguish it from the classical BKT model [5, 10, 11]. The BDT model consists of three parameters: the priors ( $P$ ), transition matrix ( $T$ ) and emission matrix ( $E$ ). The likelihood function of BDT model is given in Eq. (1) [12]: (Fig. 2)

$$P(Y_t|S_t) = \left( P(S_0|)P(Y_0|S_0) \prod_{i=1}^t P(S_i|)P(Y_i|S_i) \right) / P(Y_{0:t}|S_{0:t}) \tag{1}$$

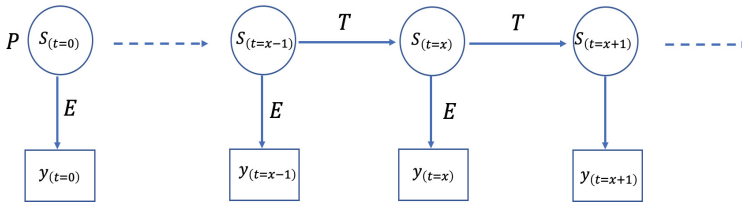


Fig. 2. HMM representation of the Bayesian diagnosis tracing model.

#### 3.1 Two-State Latent Factor Model

The BKT model does not allow for forgetting. However, such specification performs poorly in this dataset. Therefore, the BKT model reported in this paper has a full transition matrix. For the sake of comparison, the BDT parameters are reformatted in the form of BKT by ignoring the intermediate state. Table 2 shows the two models have very similar parameters. The out-sample AUC of two models are both around 0.943. In the simplest latent structure, the two models are essentially equivalent.

Table 2. Parameter comparison in the forms of BKT model

	Guess	Slip	Learn rate	Forget rate	Prior density
BDT	0.45	0.05	0.1	0.01	0.9
BKT	0.51	0.04	0.09	0.01	0.87

Table 3 reports the BDT emission probabilities. The mastery students do not skip or incur the two misconception (general misconception and miss the digit of tens).

**Table 3.** Emission probabilities of two-state BDT model

	Right	No diagnosis	Skip	Miss one	Forget borrowing	General misconception	Miss the digit of tens
No mastery	0.45	0.29	0.14	0.03	0.04	0.03	0.01
Mastery	0.95	0.03	<0.01	0.01	0.01	<0.01	<0.01

### 3.2 Three-State Latent Factor Model

This section employs a three-state model (No Mastery, Intermediate, Mastery) to better illustrate the benefit of misconception as observation. For better parameter convergence, the latent factor can only transit to the adjacent state. For the theoretical motivation of such specification, see Chap. 1 of Feng [4].

Table 4 reports the emission probabilities. The factors of the BDT model are more interpretable compared with the BKT: The no mastery state skips a lot; the intermediate state is prone to various misconceptions; the mastery state performs almost perfectly except for the most common misconceptions. The interpretable states are not only easy to communicate but also are helpful in constructing remedial instruction. In this case, students who skip and students who slip shall be treated differently: The no mastery students may need heavy intervention, such as interactive course or video tutoring; while the intermediate students can receive light-weight help, such as hint or more practices.

**Table 4.** Emission probabilities of the three-state model

Parameter	No mastery		Intermediate		Mastery	
	BDT	BKT	BDT	BKT	BDT	BKT
Right	0.32	0.11	0.63	0.79	0.96	0.97
Wrong	–	0.89	–	0.21	–	0.03
No diagnosis	0.01	–	0.24	–	0.02	–
Skip	0.67	–	0.01	–	<0.01	–
Miss one	<0.01	–	0.03	–	<0.01	–
Forget borrowing a ten	<0.01	–	0.04	–	0.01	–
General misconception	<0.01	–	0.03	–	<0.01	–
Miss the digit of tens	<0.01	–	0.01	–	<0.01	–

Besides the gain of interpretability, the BDT model also performs better in out-sample predictability. The out-sample AUC of the BDT model is 0.9243 while that of the BKT model is 0.9038.

## 4 Discussion

This paper explores the benefit of using procedure misconceptions as observation in the HMM model. The BDT model is more accurate in prediction and more interpretable in diagnosis for high dimension latent state model, when compared with the BKT model.

However, there is more work to be done. For one thing, little is known about the tail of the distribution, whose diagnosis can improve BDT performance. For another thing, the BDT model has great potential in analyzing problems that has multiple knowledge components because identified misconceptions can accurately find the component(s) to blame.

## References

1. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. *User Model. User Adapt. Interact.* **4**(4), 253–278 (1994)
2. Pelánek, R.: Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Model. User Adapt. Interact.* **27**(3–5), 313–350 (2017)
3. Desmarais, M.C., Baker, R.S.: A review of recent advances in learner and skill modeling in intelligent learning environments. *User Model. User Adapt. Interact.* **22**(1–2), 9–38 (2012)
4. Feng, J.: Essays on learning through practice. Doctoral dissertation, The University of Chicago (2017)
5. Yudelson, M.V., Koedinger, K.R., Gordon, G.J.: Individualized Bayesian knowledge tracing models. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) *AIED 2013. LNCS (LNAI)*, vol. 7926, pp. 171–180. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-39112-5\\_18](https://doi.org/10.1007/978-3-642-39112-5_18)
6. Pardos, Z.A., Heffernan, N.T.: Modeling individualization in a Bayesian networks implementation of knowledge tracing. In: De Bra, P., Kobsa, A., Chin, D. (eds.) *UMAP 2010. LNCS*, vol. 6075, pp. 255–266. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-13470-8\\_24](https://doi.org/10.1007/978-3-642-13470-8_24)
7. Piech, C., et al.: Deep knowledge tracing. In: *Advances in Neural Information Processing Systems*, pp. 505–513 (2015)
8. VanLehn, K.: *Mind Bugs: The Origins of Procedural Misconceptions*. MIT Press, Cambridge (1990)
9. Liu, R., Patel, R., Koedinger, K.R.: Modeling common misconceptions in learning process data. In: *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, pp. 369–377. ACM (2016)
10. Piech, C., et al.: Deep knowledge tracing. In: *Advances in Neural Information Processing Systems*, pp. 505–513 (2015)
11. Käser, T., Klingler, S., Schwing, A.G., Gross, M.: Beyond knowledge tracing: modeling skill topologies with Bayesian networks. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) *ITS 2014. LNCS*, vol. 8474, pp. 188–198. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-07221-0\\_23](https://doi.org/10.1007/978-3-319-07221-0_23)
12. Ghahramani, Z.: An introduction to hidden Markov models and Bayesian networks. In: *Hidden Markov Models: Applications in Computer Vision*, pp. 9–41 (2001)