Exploring Temporality and Learning in Adaptive Educational Games

Josine Verhagen, PhD, Kidaptive
jverhagen@kidaptive.com

Bayesian item response theory models for adaptive games

An educational game can be adaptive in different ways. At the extreme is real-time adaptivity, where the game adapts in difficulty every time a learner completes a challenge, with the goal to optimize engagement and learning. To achieve the continuous assessment required for real-time adaptivity, we use item response theory (IRT) models [3] from computerized adaptive testing [4, 5]. IRT provides the tools to estimate item (i.e., challenge) difficulty and person (i.e., learner) ability on a single scale.

Games differ in important ways from tests, however, especially the games for preschoolers Kidaptive is focusing on. There is often too little information to yield an accurate assessment of a learner's ability within one gameplay session, but taking all measurements together to estimate ability leaves no room to monitor growth. Numerous approaches address these concerns in computerized adaptive learning [6, 7, 8]. One approach is to use Bayesian IRT [1, 4], where prior knowledge about a person's ability is considered in the estimation process. In this way, previous gameplay results can be taken into account when learners begin a subsequent play session. The aim is to combine the previous ability estimate with the new information in such a way that the previous estimate will have more weight if we are more certain about this score, and in which the weight of the previous score decreases when more time has passed. In this way, we counter the effect of short test length by incorporating additional information, while still allowing measures of ability to change over time.

Drawing inspiration from models used in learning systems, signal processing, econometric forecasting and memory and decision making, we came up with a model which combines these properties in a natural way. This model assumes that all relevant information about previous gameplay is contained in the posterior distribution of the previous ability estimate. Subsequently, this previous gameplay information is given a weight, which decreases as a function of time passed since the last gameplay. This information can subsequently be combined with other information about the player, for example the player's age, to form a prior for the next moment of game play. Figure 1 shows the updated ability estimates after each subsequently played game item for five players of different ages. While most players are getting better by playing more, the youngest player does not seem to improve much, while the oldest player seems to have hit a ceiling right away.

CUSUM models for change in pronunciation scores

Hodoo English provides a virtual language learning world for (currently only Korean) children to increase their confidence in speaking English. Children create avatars to walk around in the virtual world and complete "Story quests" by speaking to different characters in the world. A speech recognition engine assesses the quality of their pronunciation and is used to indicate which words the player should work on. Throughout the game players repeat many of the same words and phrases. As a result, it should be possible to get a sense of their improvement in pronunciation. However, our current speech recognition engine is very sensitive to background noise and correct use of the microphone. This has often caused very low pronunciation scores even when a learner pronounced a word perfectly.

To provide useful feedback about pronunciation despite these problems, the first task is to look for indicators of erroneous or inaccurate scores: outliers in a speaker's time series of scores, or low scores which are surrounded by atypically low scores for a player, indicating background noise or a badly adjusted microphone. To develop an algorithm for this, we coded a large number of wave files as providing accurate or inaccurate scores and tried to predict inaccuracy as well as possible. We so far found a combination of features, which predicts the inaccuracy of scores correctly for about 72% of scores.

After excluding inaccurate scores, we recently started using a CUSUM method [2] to detect words on which users showed noticeable improvement. By coding improvement for a large number of files of children repeating the same word over time, we had a labeled set of improved and not improved words. From these words, we inferred the average change in speech scores indicating improvement in word pronunciation. By comparing the likelihood of a time series of a learner's scores under a model of change to the likelihood under a model of no change, and using a threshold to decide whether or not there was noticeable improvement, we aimed to identify improved and non- improved words. Event though we got good results on our original test set (75% accuracy), in the end the scores from the speech recognition engine proved to be too unreliable to use this procedure in a reliable way to detect improved scores.

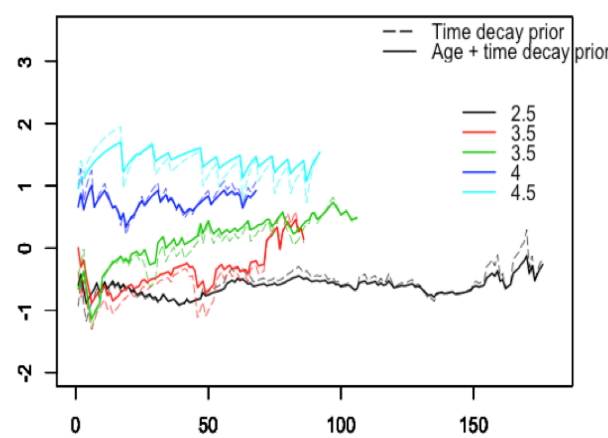

Figure 1. Estimated ability scores over time for 5 learners from different age groups (2.5 - 4.5) over time, with and without taking age into account.

REFERENCES

[1] Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media.

[2] Page, E.S. (1954). Continuous inspection schemes. Biometrika, 41, 100-115.

[3] Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum.

[4] van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. New York, NY: Springer

[5] Wainer, H., Dorans, N. J., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., & Steinberg, L. (2000). *Computerized adaptive testing: A primer*. Mahwah, NJ: Erlbaum.

[6] Eggen, T. J. (2012). Computerized adaptive testing item selection in computerized adaptive learning systems. *Psychometrics in practice at RCEC.* Retrieved from http://doc.utwente.nl/80211/1/Chapter_2.pdf

[7] Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. J. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers and Education,* 57, 1813–1824.

[8] Wauters, K., Desmet, P., & van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: possibilities and challenges. *Journal of Computer Assisted Learning*, 26, 549–562