

# Project template

**Matjaž Pogačnik**

faculty of computer and information science  
večna pot 113  
1000 ljubljana

**Abstract.** Kratek opis ideje

**Key words:** equiz, rating

## 1 INTRODUCTION

Ena izmed funkcionalnosti aplikacije eQuiz je rating študentov in nalog preko Elo rating sistema. Študent se sooči z nalogo in na ta način se iztočasno izračuna rating študenta in rating naloge, kot bi igrala študent in naloga šah. Študent lahko na nalogo odgovori prav ali napačno, kar je ekvivalentno temu, da študent naloge premaga ali proti njej izgubi. Pojavijo pa se pomankljivosti takega rating sistema.

Če uporabljamo eQuiz za spremljanje ratinga študentov in nalog v sklopu določenega predmeta, bo proti številu študentov, ki bodo v tem letu aktivni, število nalog navadno mnogo večje. To pa pomeni, da bodo, dokler študentje ne rešujejo samih nalog, ti prejeli posodobitev ratinga veliko večkrat kot kot posamezna naloga. Še več posodobitev pa študentje, ki so pri reševanju nalog bolj aktivni. Torej ratingini, pripisani določeni nalogi ali študentu lahko izvirajo iz več ali manj reševanj, kar naredi rating bolj ali manj verodostojen. Zaupanje v določen rating lahko tako iz posamezne številke raje razširimo na interval zaupanja.

## 2 GLICKO RATING

Za tak pristop je primeren Glicko rating, ki poleg ratinga za posameznega igralca (študent ali naloga) vpeljuje še deviacijo  $RD$ , ki nam omogoča predstavitev ratinga posameznega igralca kot 95% interval zaupanja:

$$(r - 1.96RD, r + 1.96RD) \quad (1)$$

pri čemer se  $RD$  manjša ob vsakem updatu ratinga, kjer se igralec sooči z drugim igralcem (ali več njih)

$$RD' = \sqrt{\left(\frac{1}{RD^2} + \frac{1}{d^2}\right)^{-1}} \quad (2)$$

kjer  $RD$  predstavlja deviacijo pred posodobitvijo,  $d^2$  pa je definiran kot

$$d^2 = \left( q^2 \sum_{j=1}^m (g(RD_j))^2 E(s|r, r_j, RD_j) (1 - E(s|r, r_j, RD_j)) \right)^{-1} \quad (3)$$

$$q = \frac{\ln 10}{400} = 0.0057565 \quad (4)$$

$$E(s|r, r_j, RD_j) = \frac{1}{1 + 10^{-g(RD_j)(r-r_j)/400}} \quad (5)$$

$$g(RD) = \frac{1}{\sqrt{1 + 3q^2(RD^2)/\pi^2}} \quad (6)$$

Nov rating  $r'$  se izračuna po formuli

$$r' = r + \frac{q}{1/RD^2 + 1/d^2} \sum_{j=1}^m g(RD_j) (s_j - E(s|r, r_j, RD_j)) \quad (7)$$

Zgornje formule so posplošene za posodobitev ratinga  $r$  in deviacije  $RD$  igralca proti skupini  $m$  nasprotnikov z deviacijami  $RD_1, RD_2, \dots, RD_m$  in ratingi  $r_1, r_2, \dots, r_m$ .  $s_1, s_2, \dots, s_m$  predstavljajo izide, ki so lahko 0 ali 1, za izgubo ali zmago.

Igralčev  $RD$  se ne posodablja samo, ko se ta sooča, temveč tudi ob preteku določene časovne periode, kar predstavlja zniževanje verodostojnosti trenutnega ratinga, če igralec določen čas ne igra. Tako se njegov  $RD$  posodobi kot

$$RD = \min \left( \sqrt{RD_{old}^2 + c^2}, 350 \right) \quad (8)$$

konstanta  $c$  je, poleg maksimalne deviacije, ki je v tem primeru 350, edini parameter, ki ga nastavlja administrator sistema. Več o tem v kontekstu equiza v poglavju X

Matematične izpeljave formul so na voljo na tukaj

Iz formul vidimo, da je  $r'$  odvisen od  $RD$  igralca in ni uravnotežen tako, kot je pri Elo sistemu, kjer se zmagovalcu zviša rating toliko, kot se poražencu zniža. Ker naj bi velikost  $RD$  odražala koliko informacije imamo o igralcu, se v iteraciji ocenjevanja, kjer ima en igralec velik  $RD$ , drug pa majhen, prvemu zviša veliko več, kot se drugemu zniža, saj vemo, da je rating prvega nezanesljiv, rating drugega pa je. Enako se v tem primeru ne more upravičeno znižati rating drugega igralca toliko, kot bi se, če bi bil rating njegovega soigralca zanesljiv.

A useful way to summarize a player's strength is through a confidence interval (or more particularly, given the quasi-Bayesian derivation of the Glicko system, a "credible" interval) rather than just reporting a rating. The confidence interval has the interpretation of reporting the interval of plausible values for the player's actual strength, acknowledging that a rating is merely an estimate of a player's unknown true strength. A common choice is to report a 95% confidence interval which provides 95% confidence that the player's true ability is within the interval. The formula for a 95% confidence interval for a player with rating  $r$  and ratings deviation  $RD$  is given by

### 3 IZRAČUNI IZ EQUIZ PODATKOV

#### 3.1 Pridobljeni podatki

Iz podatkovne baze eQuiza so bili pridobljeni trenutni ratingi študentov in nalog ter kronološki zapis reševanj quizov, sestavljenih iz različnih nalog. Podatek o reševanju posamezne naloge v quizu je bil izračunan na podlagi sprememb ratingov nalog po končanem quizu. V tabelo podatkov je bil tako dodan stolpec correct z vrednostjo 0 ali 1 za vsako nalogo, kjer so naloge s podvprašanji razdeljene v posamezne naloge. Ta podatek je bil uporabljen tudi v kasnejšnjih izračunih ELO in Glicko ratingov.

Za izračun Glicko ratinga je, za razliko od ELO ratinga, potreben še čas reševanja. Zato so bile posodobitve ratingov združene s tabelo reševanj, ki vsebuje potrebne čase. Pri tem so se pojavila neskladja, kjer nekatere posodobitve ratingov, ki se nanašajo na nek identifikator kviza, niso bile nikdar zabeležene, nekatera zabeležena reševanja pa se ne nanašajo na noben rating, kar pomeni, da so k izračunu ratinga nekoč lahko prispevali različni tipi ocenjevanj, trenutno pa se v rating štejejo samo Rating Quizi. Tako so bili pridobljeni podatki za 305 različnih uporabnikov in 1348 nalog, za nadaljnje izračune pa so bili uporabni podatki za 261 uporabnikov in 1299 nalog.

Za izračune ratingov so bile od najstarejšega ratinga uporabnika po času reševanj zaporedoma pregledane naloge iz ustreznega kviza, ki so bile glede na prejšnji

najstarejši ali (osnovni rating) uporabnika in naloge ponovno ratane po izbranem algoritmu. Tako je bil po zaporednih nalogah glede na reševanje (stolpec correct) izračunan nov rating uporabnika in novi ratingi nalog, ki so bili sproti zapisani v tabele na mesta z ustreznimi časi in identifikatorji izpitov, za uporabo v naslednjih korakih algoritma. Ko smo dosegli konec zabeleženih ratingov smo imeli izračunane ratinge za vse naloge in userje.

#### 3.2 Trenutni rating sistem - ELO

Pri oddaji kviza so pridobljeni prejšnji rating posameznih nalog v kvizu in prejšnji rating uporabnika. Za naloge in uporabnika so potem zaporedoma po nalogah izračunani novi ratingi po naslednjih enačbah:

$$R'_A = R_A + K \cdot (S_A - E_A) \quad (9)$$

$$E_A = \frac{Q_A}{Q_A + Q_B} \quad (10)$$

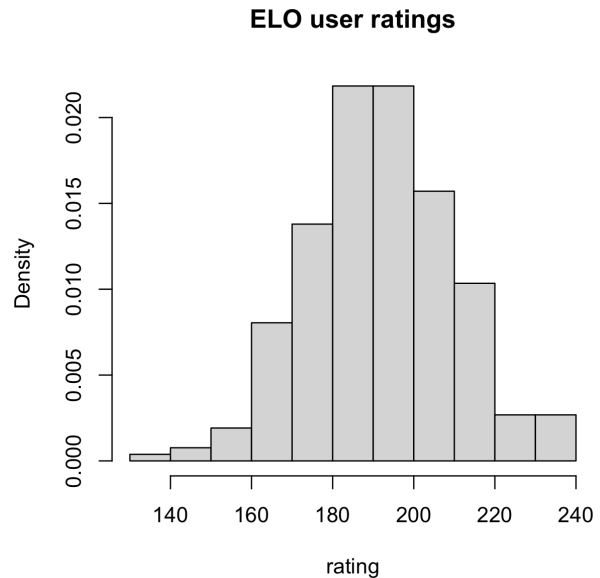
Kjer so

$$Q_A = 10^{R_A/66} \quad Q_B = 10^{R_B/66} \quad (11)$$

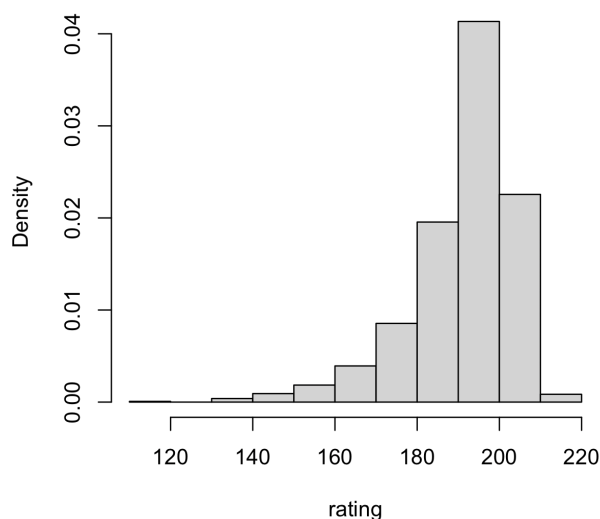
$S_A$  predstavlja dejanski izid, v primeru eQuiza 0 ali 1,  $E_A$  pa pričakovan izid.

$K$  je po navadi izračunan glede na število iger na katerih temelji trenutni rating (plus število iger v turnamentu pri šahu), v trenutni implementaciji ratinga pa je privzet kot konstanten z vrednostjo 4,5

Po takem postopku je bil po zgoraj opisanih algoritmih izračunan ELO rating na filtriranih podatkih z osnovnim ratingom 200.



ELO exercise ratings



### 3.3 Glicko rating na pridobljenih podatkih

S tem pridobimo v primeru študentov informacijo o sprotnem delu, saj bo RD manjši za učence, ki so pogostejše ocenjeni (poleg opravljenih sprotnih preverjanj še naključna reševanja kvizov/nalog), kar kaže na več/manj sprotnega dela. (Glede na spreminjanje RD lahko spremljamo delo učenca kot nekakšen koeficient napredka/ak.vnos.) Za nasprotnika učencu – nalogo, pa deviacija predstavlja kdaj je bila naloga nazadnje ocenjena, torej koliko je njena ocena zanesljiva. V primeru učenca, ki izbira naloge iz zbirke na equizu, pa lahko naloge z veliko deviacijo iden.ficiramo kot nepriljubljene (malo učencev se lo. naloge). Reciklirane naloge bodo torej tudi večkrat nastopile na preverjanjih in bo natančnost njenih ra.ngov večja (dobro za parametrizirane naloge, ker se zaradi njihove narave lahko večkrat pojavijo v preverjanjih) Ra.ng nalog, ki se pojavijo na izpitu, pa bo še vedno zanesljiv, saj bo v is. iteraciji nalogo reševalo veliko učencev. Ra.ng učenca/naloge, ki je neak.ven čez določen časovni interval se ustrezno zmanjša, kar je poleg maksimalne deviacije edini parameter, ki ga je potrebno nastaviti kot administrator. Kako sistematično določamo ta parameter je opisano v dodatnem članku o glicko, vendar ga lahko verjetno vezemo na redna preverjanja kot npr. kolokvije. Ker se da predmet opravi. samo s kolokviji celotno snov sta.s.ke razdelijo na 5 delov. Torej bo študent, ki opravi samo eno preverjanje od pe.h ustrezno imel manjši ra.ng predvsem pa večjo deviacijo (če je opravljal samo prvega se bo njegov RD š.rikat zviševal do konca semestra), saj imamo informacijo o njegovem znanju čedalje bolj nezanesljivo, ker vemo čedalje manj koliko študent dejansko ve o sta.s.ki odkar je reševal .s. kolokvij (pričakuje se, da študent tekom leta ve več in več o

sta.s.ki).

*Lemma 3.1.*  $\ell = \varepsilon + t(p-1)(q-1)$ .

*Proof.*

$$\begin{aligned}\ell &= \varepsilon + \mathbf{r}(q-1) \\ &= \varepsilon + t(p-1)(q-1)\end{aligned}$$

*Theorem 3.2.*  $\ell \equiv \varepsilon \pmod{\varphi(n)}$ .

*Proof.* By Lemma 3.1 and (??) we have  $\varepsilon \bmod \varphi(n) = \ell$ . ■

*Corollary 3.3.*  $\exists k \in \mathbb{Z} : \ell - \varepsilon = k\varphi(n)$ . ■

*Proof.* Follows by Theorem 3.2. ■

## 4 GLICKO ZA SPREMLJANJE ŠTUDENTOV

### 4.1 Implementacija

## 5 CONCLUSION

conclusion