

КЛАСТЕРИЗАЦИЯ ДИАГНОСТИЧЕСКИХ ТЕСТОВ ПРИ ИЗУЧЕНИИ ПРЕДЛОГОВ АНГЛИЙСКОГО ЯЗЫКА В СООТВЕТСТВИИ С ТАКСОНОМИЕЙ БЛУМА

© 2015 С.А. Прохоров, С.А. Сучкова, И.М. Куликовских

Самарский государственный аэрокосмический университет имени академика С.П. Королева
(национальный исследовательский университет)

Поступила в редакцию 30.07.2015

Данная работа посвящена вопросам кластеризации банков диагностических тестов при изучении предлогов английского языка в рамках адаптивного обучения. В основу разработанной методики кластеризации положена таксономия Блума, что предполагает формирование банка заданий в соответствии с когнитивными уровнями. Для представления методики кластеризации и алгоритмов «multiple choice» введен ряд определений и обозначений. Проведена серия вычислительных экспериментов на основе смоделированных ответов для проверки адекватности предложенной методики с помощью методов k-means, c-means и иерархической кластеризации.

Ключевые слова: предлог, таксономия Блума, адаптивное обучение, машинное обучение, кластеризация, когнитивный уровень, диагностический тест

ПОСТАНОВКА ЗАДАЧИ

Использование предлогов является одной из самых сложных областей для изучения английского языка как иностранного [1, 2]. Так, например, авторы [3] отмечают трудности в использовании предлогов учащимися, что подтверждает анализ корпуса ошибок, в котором предлоги составляют 12%. Это связано с особенностями английского предлога как части речи. В среднеанглийском периоде возросла грамматическая роль предлога при отмирании падежной системы. Теперь предлоги играют важную роль в построении английского предложения; они отражают пространственные, временные, причинные и иные отношения между значимыми словами.

С течением времени многие предлоги частично утратили свое первичное семантическое значение [4] и теперь обладают многозначностью. Так, зарегистрировано 18 основных использований предлога *at* [2]. В то же время разные предлоги могут передавать одинаковые отношения. Это особенно касается наиболее распространенных предлогов, таких как *in*, *of*, *to*, которые относятся к самым частотным словам в английском языке [1]. Следует упомянуть значительные сложности, связанные с омонимичностью предлогов [5-8], например, места и времени [7]. Общая тенденция описания данного феномена определяется

перечислением возможных значений с поверхностным объяснением связи между пространственным значением и непространственными расширениями [5]. Исключением данной тенденции является работа [8], которая для представления предлогов использует когнитивный подход, частично опираясь на теорию метафоры. Как отмечают авторы [5], несмотря на недостатки, связанные с отсутствием анализа концептуальной теории метафоры и полисемии, а также привязыванием метафорических значений к ряду предлогов без достаточных объяснений, использование полисемической модели [9,10] позволяет значительно улучшить подход к изучению языка.

Другая потенциальная трудность связана с тем, что однокоренные слова, принадлежащие к разным частям речи, и слова со связанными или близкими значениями зачастую требуют различных предлогов [1]. Дополнительную трудность создает также и тот факт, что наблюдаются случаи сращения предлога со связываемой им знаменательной частью речи в единое смысловое целое, например *at night*, *angry with*, *look for* [4]. Употребление предлога в данных сочетаниях часто немотивировано и требует механического запоминания.

Анализируя английский как иностранный, описанные трудности являются дополнением к потенциальной неопределенности, вызванной отсутствием связи между системами предлогов в родном и иностранном языках [1,5]. В частности, такие кросс-культурные вариации включают разницу в типе грамматической формы и сегментации пространственных сцен при использовании предлогов места [5]. Все это делает предлоги чрезвычайно сложными в обучении и преподавании. Учитывая сложность данного языкового явления, была поставлена задача создания адекватной модели, которая бы сформировала грамматическую основу

Прохоров Сергей Антонович, доктор технических наук, профессор, заведующий кафедрой информационных систем и технологий. E-mail: sp.prokhorov@gmail.com

Сучкова Светлана Анатольевна, кандидат филологических наук, доцент кафедры иностранных языков. E-mail: s.suchkova@gmail.com

Куликовских Илона Марковна, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: kulikovskikh.i@gmail.com

для представления предлогов и способствовала бы его корректному выбору в новом контексте.

Одним из решений данной проблемы является применение адаптивного метода обучения, который основан на корректировке обучающего контента и способов его представления, исходя из действий обучаемого. В качестве основы предлагается использовать таксономию [11-15], созданную Б. Блумом в 1956 [11] и впоследствии доработанную Л. Андерсоном [12]. В этой теории уровни мышления описываются по возрастанию сложности задачи с помощью следующих глаголов: помнить, понимать, применять, анализировать, оценивать, создавать. Каждый уровень представляет собой набор мыслительных операций для решения определенной образовательной задачи. В этом смысле, таксономия Блума может являться основой при формировании образовательного контента, создания заданий и их оценивания. Применяя данную теорию к проблеме изучения предлогов, необходимо на каждом этапе предлагать материал, соответствующий заданному когнитивному уровню для повышения эффективности обучения.

Неотъемлемой частью адаптивного обучения является диагностическое тестирование, позволяющее оценить текущий уровень знаний в изучаемой области. Прежде чем определить стартовый уровень освоения материала, необходимо разработать банки диагностических тестов [16-19], что является трудоемкой задачей ввиду необходимости соответствия контента строго определенному когнитивному уровню, учета схожести вопросов в рамках одного банка заданий и последовательности их представления согласно таксономии. В целях упрощения данной задачи, было сделано множество попыток автоматизировать этот процесс [17-19] с помощью алгоритмов машинного обучения в рамках создания систем адаптивного обучения [20,21].

Важным аспектом, который необходимо учитывать при тестировании, является инструмент анализа, который определяет тип вопросов. В работе [21] автор указывает на «multiple choice» как на универсальный инструмент для формирования пары «вопрос-ответ» даже на высоких когнитивных уровнях. Данная точка зрения в последующем получила подтверждение в работах [22-25], однако, требует проведения дополнительного анализа случайности выбора правильных ответов с учетом психологических факторов [26-28].

Таким образом, целью данной работы является кластеризация банка диагностических тестов при изучении предлогов английского языка в соответствии с таксономией Блума.

ФОРМАЛИЗАЦИЯ ЗАДАЧИ

Анализ проведенных исследований в области адаптивного обучения показал, что существует множе-

ство подходов для решения поставленной задачи [29-35], среди которых используются алгоритмы машинного обучения без учителя (кластеризации). Данные алгоритмы позволяют разбить множество объектов на группы схожих, сократить объем данных, оставив по одному типичному представителю, а также выделить нетипичные объекты, неподходящие ни одному из кластеров. Ввиду известного значения числа кластеров I , заданного числом когнитивных уровней таксономии, наиболее целесообразным является применение следующих методов [36-40]:

- 1) k-means с четким разбиением на уровни;
- 2) c-means с определением степени принадлежности через задание коэффициента нечеткости;
- 3) иерархическая кластеризация, имитирующая структуру таксономии Блума.

Для представления методики кластеризации банка диагностических тестов в соответствии с таксономией Блума, введем ряд определений.

Определение 1. Пусть U – конечное множество ответов обучаемых, X – конечное множество когнитивных уровней. Тогда отображение $x_0 : U \rightarrow X$ называется задачей диагностического тестирования.

Определение 2. Отображение $r : U \rightarrow D_r$, где $D_r = \{0, 1\}$, называется корректностью ответа.

Определение 3. Пусть $\{q_m\}_{m=1, M}$ – M банков с N тестовых заданий $q_m = \{t_{m,n}\}_{n=1, N}$ в каждом. Тогда вектор $u^r = (r(q_1), \dots, r(q_m))$ называется результатами оценки ответов $u \in U$ на вопросы $q \in Q$.

Определение 4. Пусть $q^x \in Q$ – банки тестов, соответствующие когнитивным уровням $x \in X$.

Тогда отображение $q^x : Q \rightarrow X$ называется задачей кластеризации банка вопросов Q с мерой $\rho(u, u')$.

С учетом приведенной формализации, представим методику кластеризации банка тестов в следующем форме:

- 1) формирование набора тестовых заданий $\{t_{m,n}\}_{m=1, M, n=1, N}$ для каждого банка $q_m \in Q$;
- 2) задание метрического пространства в виде результатов оценки ответов $\bar{u}^r \in [0, 1]$ в пределах одного банка на основе всей совокупности $u \in U$;
- 3) выбор меры $\rho(\bar{u}, \bar{u}')$ в метрическом пространстве $\bar{u}^r \in [0, 1]$;
- 4) оценивание косинусных мер сходства векторов

$(r(t_{m,1}), \dots, r(t_{m,n})) \in D_r, n = 1..N$ для каждого

банка $q_m \in Q$ с построением симметричной матрицы сходства между парами векторов через нахождение ска-

лярного произведения $r(t_{m,i}) \cdot r(t_{m,j}), i, j = 1..N$;

5) реализация алгоритмов кластеризации значе-

ний $\bar{u}^r \in [0, 1]$ с мерой $\rho(\bar{u}, \bar{u}')$.

Сформируем банки диагностических тестов

$\{q_m\}_{m=1..M}$, где $M = 9$, и зададим алгоритмы функционирования инструмента «multiple choice» для

формирования вектора $(r(t_{m,1}), \dots, r(t_{m,n})) \in D_r,$

$n = 1..N$ для каждого банка $q_m \in Q$ в виде таблицы 1, где для компактности представления введены следующие обозначения:

L – возможное количество вариантов ответа, предоставляемых «multiple choice»;

P – множество всех предлогов;

O – множество всех предлогов, обладающих омонимичностью $O \subset P$;

C – множество всех предлогов, имеющих коллокации с заданным словом $C \subset P$;

O_ϕ – множество всех предлогов, обладающих омонимичностью, с пространственными значениями

$\phi: P \rightarrow O_\phi$;

O_ψ – множество всех предлогов, обладающих омонимичностью, с метафорическими значениями

$\psi: P \rightarrow O_\psi$.

Таблица 1. Банки диагностических тестов и алгоритмы анализа ответов

	Инструкция	Алгоритм «multiple choice»
1	Choose the correct preposition in brackets, if necessary.	$N = 5, L = 3$ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
2	In the text field below, write the numbers of the sentences in which there is no preposition.	$N = 5, L = 5$ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
3	Choose appropriate prepositions from the list below to complete the sentences. in, at, on, of, on, -	$N = 5, L = 6$ $(\{u_i^r u_i^r = 0\}_n), l = 1..L, n = 1..N$
4	Complete the sentences with correct prepositions if necessary or write '-' instead.	$N = 5, L = P $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
5	Correct mistakes in prepositions if necessary. Write the correct preposition or '-' instead.	$N = 5, L = P $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
6	Fill in the blanks using the best choice(s) from the suggested lists.	$N = 5, L = C $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
7	Reorder the sentences on your right so that they have the same preposition in each line with the sentences on the left. Write these prepositions in the text field below one by one separating them by commas.	$N = 5, L = O $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
8	What picture best describes the sentence 'The cat jumped over the wall.'? Write the appropriate letter(s) in the text field below.	$N = 4, L = O_\phi $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$
9	Read the text below. Paraphrase each preposition OVER, using a phrase or a word, which is similar in meaning. Write these phrases or words in the text field, separating them by commas.	$N = 4, L = O_\psi $ $(\{u_i^r u_i^r = 0\}_n), n = 1..N, l = 1..L$

ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

С целью проверки адекватности предложенной методики, была проведена серия вычислительных экспериментов для каждого из перечисленных выше методов кластеризации, чтобы повысить качество полученных результатов. Ниже представлен код, реализованный в пакете R.

```
install.packages («el071»)
install.packages («lsa»)
# инициализация
n<-10; N1<-5; N2<-4; t<-0:1
# моделирование ответов
matrix_t1<-c(); matrix_t2<-c();
matrix_t3<-c(); matrix_t4<-c(); ma-
trix_t5<-c()
matrix_t6<-c(); matrix_t7<-c(); ma-
trix_t8<-c(); matrix_t9<-c()
for(i in 1:N1){
  matrix_t1 <- cbind(matrix_t1,
sample(t, n, replace = TRUE))
  matrix_t2 <- cbind(matrix_t2,
sample(t, n, replace = TRUE))
  matrix_t3 <- cbind(matrix_t3,
sample(t, n, replace = TRUE))
  matrix_t4 <- cbind(matrix_t4,
sample(t, n, replace = TRUE))
  matrix_t5 <- cbind(matrix_t5,
sample(t, n, replace = TRUE))
  matrix_t6 <- cbind(matrix_t6,
sample(t, n, replace = TRUE))
  matrix_t7 <- cbind(matrix_t7,
sample(t, n, replace = TRUE))
}
for(i in 1:N2){
  matrix_t8 <- cbind(matrix_t8,
sample(t, n, replace = TRUE))
  matrix_t9 <- cbind(matrix_t9,
sample(t, n, replace = TRUE))
}
# меры сходства
t1_sim<-cosine(matrix_t1); t2_sim<-
cosine(matrix_t2)
t3_sim<-cosine(matrix_t3); t4_sim<-
cosine(matrix_t4)
t5_sim<-cosine(matrix_t5); t6_sim<-
cosine(matrix_t6)
t7_sim<-cosine(matrix_t7); t8_
sim<-cosine(matrix_t8); t9_sim<-
cosine(matrix_t9)
# матрица ответов в метрическом
пространстве
vec_q1<-c(); vec_q2<-c(); vec_q3<-
c(); vec_q4<-c(); vec_q5<-c()
vec_q6<-c(); vec_q7<-c(); vec_q8<-
c(); vec_q9<-c(); matrix_q<-c()
for(i in 1:n){
  vec_q1 <- c(vec_q1, sum(matrix_
```

```
t1[i,])/N1)
  vec_q2 <- c(vec_q2, sum(matrix_
t2[i,])/N1)
  vec_q3 <- c(vec_q3, sum(matrix_
t3[i,])/N1)
  vec_q4 <- c(vec_q4, sum(matrix_
t4[i,])/N1)
  vec_q5 <- c(vec_q5, sum(matrix_
t5[i,])/N1)
  vec_q6 <- c(vec_q6, sum(matrix_
t6[i,])/N1)
  vec_q7 <- c(vec_q7, sum(matrix_
t7[i,])/N1)
  vec_q8 <- c(vec_q8, sum(matrix_
t8[i,])/N2)
  vec_q9 <- c(vec_q9, sum(matrix_
t9[i,])/N2)
}
matrix_q<-cbind(vec_q1,vec_q2,vec_
q3,vec_q4,vec_q5,vec_q6,vec_q7,vec_
q8,vec_q9)
dtest <-data.frame(tb1 = vec_q1, tb2
= vec_q2, tb3 = vec_q3, tb4 = vec_q4,
tb5 =vec_q5, tb6 = vec_q6, tb7 = vec_q7,
tb8 = vec_q8, tb9 = vec_q9)
# задание меры
dtest <-apply(dtest, 2, diff); dtest
<-t(dtest)
# k-means
dtest_km<- kmeans(dtest, 6, 10000);
print(dtest_km)
# c-means
dtest_cm<-cmeans(dtest, 6, 10000,
verbose=TRUE,method=»cmeans»,m=6);
print(dtest_cm)
# иерархическая кластеризация
dtest_hc <-hclust(dist(dtest));
print(dtest_hc)
plot(dtest_hc)
rect.hclust(dtest_hc, k=6)
groups <- cutree(dtest_hc, k=6)
```

Приведем основные результаты проведенных экспериментов. Смоделированные ответы на тестовые вопросы в метрическом пространстве представлены в таблице 2. В таблице 3 – результаты работы алгоритмов k-means и c-means, на рисунке 1 – дендрограмма с результатами иерархической кластеризации, построенная в пакете R.

С учетом того, что порядок уровня не привязывался к заданному диагностическому банку тестовых заданий, результаты, полученные с помощью каждого из рассмотренных уровней, можно считать идентичными.

Из полученных результатов следует, что, например, тестовый банк q_4 выделен в отдельный кластер. Ниже представлены матрицы ответов на тестовые задания в банке $\left\{r(t_{m,n})\right\}_v$, где $m = 4$, с количеством испытаний $v = 10$, $N = 5$, $n = 1$. N , и

Таблица 2. Смоделированные ответы в метрическом пространстве

	tb1	tb2	tb3	tb4	tb5	tb6	tb7	tb8	tb9
1	0,2	0,4	0,8	0,6	1,0	0,4	0,2	0,75	0,5
2	0,4	0,4	1,0	0,4	0,4	0,4	0,4	0,5	0,25
3	0,6	0,4	0,6	0,4	0,8	0,6	0,4	0,5	0,75
4	1,0	0,6	0,6	1,0	0,4	0,6	0,2	0,25	0,5
5	0,8	0,2	0,6	0,6	0,8	0,8	0,8	0,5	0,5
6	0,6	0,4	0,6	1,0	0,2	0,4	0,6	0,0	0,25
7	0,2	0,4	0,6	0,8	0,4	0,2	0,4	0,5	0,25
8	0,4	0,8	0,4	0,2	1,0	0,6	0,6	0,5	0,5
9	0,2	0,4	0,6	0,0	0,6	0,6	0,8	1,0	1,0
10	0,6	0,2	0,6	1,0	0,8	0,2	0,6	0,5	0,25

Таблица 3. Результаты кластеризация с помощью k-means и c-means

	tb1	tb2	tb3	tb4	tb5	tb6	tb7	tb8	tb9
k-means	5	5	1	3	2	4	4	6	6
c-means	5	5	1	4	3	2	2	6	6

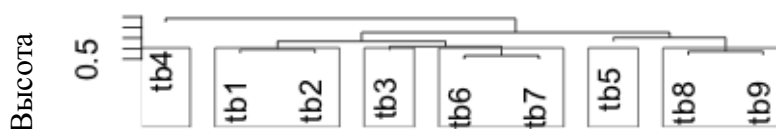


Рис. 1. Дендрограмма с иерархической кластеризацией банка диагностических тестов

матрица косинусных мер сходства между заданиями $r(t_{m,i}) \cdot r(t_{m,j})$, $i, j = 1..N$, соответственно:

$$\left(\{r(t_{m,n})\}_y \right) = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$r(t_{m,i}) \cdot r(t_{m,j}) = \begin{pmatrix} 1 & 0,833 & 0,612 & 0,722 & 0,833 \\ 0,833 & 1 & 0,612 & 0,722 & 0,833 \\ 0,612 & 0,612 & 1 & 0,707 & 0,612 \\ 0,722 & 0,722 & 0,707 & 1 & 0,866 \\ 0,833 & 0,833 & 0,612 & 0,866 & 1 \end{pmatrix}.$$

Представленные значения матрицы схожести позволяют сделать вывод об аналогичности тестовых заданий, представленных в одном банке.

ВЫВОДЫ

В рамках данной работы были сформированы банки диагностических тестов при изучении предлогов английского языка и создана методика их кластеризации на основе таксономии Блума.

С целью формального описания поставленной задачи кластеризации были сформулированы определения задач классификации и кластеризации, а также корректности ответа и результата оценки ответа на заданный тестовый вопрос. Предложенные определения позволили описать структуру алгоритма «multiple choice», который, в свою очередь, формирует матрицы ответов для определения меры сходства и задания метрического пространства, необходимого для реализации алгоритмов кластеризации методами k-means, c-means и иерархической кластеризации. Серия вычислительных экспериментов подтвердила целесообразность полученных теоретических результатов и адекватность предложенной методики.

Дальнейшее направление исследований связано с проведением тестирования на совокупности обучаемых с их последующей классификацией на когнитивные уровни и с построением оптимальных обучающих траекторий с помощью деревьев решений.

ФИНАНСИРОВАНИЕ

Работа выполнена при государственной поддержке Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013-2020 годы.

СПИСОК ЛИТЕРАТУРЫ

1. *De Felice R.* Automatic detection of preposition errors in learning writing // *CALICO Journal*. 2009. Vol. 26. №3. P. 512-528.
2. *Swan M.* Practical English Usage: Second edition. Oxford: Oxford University Press, 2003. 658 p.
3. *Parrot M.* Grammar for English language teachers: Second edition. Cambridge: Cambridge University Press, 2010. 479 p.
4. *Аксененко Б.Н.* Предлоги английского языка. Москва: Издательство литературы на иностранных языках, 1956. 320 с.
5. *Tyler A., Evans V.* The semantics of English prepositions. Cambridge: Cambridge University Press, 2003. 238 p.
6. *Kleiner L.F.* The semantics of English prepositions. Book review // *Journal of Pragmatics*. 2005. Vol. 37. P. 775-779.
7. *Kemmerer D.* The spatial and temporal meanings of English prepositions can be independently impaired // *Neuropsychologia*. 2005. Vol. 43. P. 797-806.
8. *Lindstromberg S.* English prepositions explained: Revised edition. Amsterdam: John Benjamins, 2010. 287 p.
9. The iconicity of embodied meaning. Polysemy of spatial prepositions in the cognitive framework / *F.V. Gucht, K. Willems, L. Cuypere* // *Language Sciences*. 2007. Vol. 29. P. 733-754.
10. *Mueller C.M.* English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? // *System*. 2011. Vol. 39. P. 480-490.
11. Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain / *B.S. Bloom (Ed.), M.D. Engelhart, E.J. Furst, W.H. Hill, D.R. Krathwohl*. New York: David McKay, 1956. 207 p.
12. A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives / *L.W. Anderson (Ed.), D.R. Krathwohl (Ed.), P.W. Airasian, K.A. Cruikshank, R.E. Mayer, P.R. Pintrich, J. Raths, M.C. Wittrock*. New York: Longman, 2001. 336 p.
13. Bloom's taxonomy revised: specifying assessable learning objectives in computer science / *C.W. Starr, B. Manaris, R.H. Stalvey* // The proceedings of the 39th SIGSCE Technical Symposium on Computer Science Education. 2008. P. 261-265.
14. *Лазарева И.Н.* Таксономический подход в проектировании личностно ориентированного интеллектуально-развивающего обучения // *Известия Российского государственного педагогического университета им. А.И. Герцена*. 2009. № 94. С. 130-136.
15. *Туласынова Н.Ю.* Развитие критического мышления студентов в процессе обучения иностранному языку // *Известия Российского государственного педагогического университета им. А.И. Герцена*. 2009. № 112. С. 130-136.
16. *Романов Е.В.* Разработка системы оценки учебных достижений студентов в контексте реализации компетентностного подхода // *Инновации в образовании*. 2011. № 4. С. 71-76.
17. Automated analysis of exam questions according to Bloom's taxonomy / *N. Omar, S.S. Haris, R. Hassan, H. Arshad, M. Rahmat, N.F.A. Zainal, R. Zulkifli* // *Procedia - Social and Behavioral Sciences*. 2012. Vol. 59. P. 297-303.
18. Analysing the cognitive level of classroom questions using machine learning techniques / *A.A. Yahya, A. Osman, A. Taleb, A.A. Alattab* // *Procedia - Social and Behavioral Sciences*. 2013. Vol. 97. P. 587-595.
19. *Chang W.-C., Chung M.-S.* Automatic applying Bloom's Taxonomy to classify and analysis the cognition level of English question items // The proceedings of Joint conferences on pervasive computing. 2009. P. 727-734.
20. *Hashim K., Khairuddin N.N.* Software engineering assessments and learning outcomes // The proceedings of the 8th WSEAS Int. Conference on Software Engineering, Parallel and Distributed Systems. 2009. P. 131-134.
21. *Thelwall M.* Computer-based assessment: a versatile educational tool // *Computers & Education*. 2000. Vol. 34. P. 37-49.
22. Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers / *E. Ventouras, D. Triantis, P. Tsiakas, C. Stergiopoulos* // *Computers & Education*. 2010. Vol. 54. P. 455-461.
23. A comparison between three and four option multiple choice questions / *A. Dehnad, H. Nasser, A.F. Hosseini* // *Procedia - Social and Behavioral Sciences*. 2014. Vol. 98. P. 398-403.
24. *Ghafournia N.* The relationship between using multiple-choice test-taking strategies and general language proficiency level // *Procedia - Social and Behavioral Sciences*. 2013. Vol. 70. P. 90-94.
25. *Hemmati F., Ghaderi E.* The effect of four formats of multiple-choice questions on the listening comprehension of EFL learners / *F. Hemmati*, // *Procedia - Social and Behavioral Sciences*. 2014. Vol. 98. P. 637-644.
26. *Espinosa M.P., Gardeazabal J.* Optimal correction for guessing in multiple-choice tests // *Journal of Mathematical Psychology*. 2010. Vol. 54. – P. 415-425.
27. *McCausland W.J., Marley A.A.J.* Prior distributions for random choice structures // *Journal of Mathematical Psychology*. 2014. Vol. 57. P. 78-93.
28. *McCausland W.J., Marley A.A.J.* Bayesian inference and model comparison for random choice structures // *Journal of Mathematical Psychology*. 2014. Vol. 62-63. P. 33-46.
29. Multiple instance learning for classifying students in learning management systems / *A. Zafra, C. Romero, S. Ventura* // *Expert Systems with Applications*. 2011. Vol. 38. P. 15020-15031.
30. *Paul D.V., Pawar J.D.* A grouping strategy based partition algorithm for clustering questions in a question bank // *Journal of Convergence Information Technology*. 2014. Vol. 19. № 2. P. 70-78.
31. *Thomassey S., Fiordaliso A.* A hybrid sales forecasting system based on clustering and decision trees // *Decision Support Systems*. 2006. Vol. 42. P. 408-421.
32. Automatic leveling system for e-learning examination pool using entropy-based decision tree / *S.-C. Cheng, Y.-M. Huang, J.-N. Chen, Y.-T. Lin* // The proceedings of the 4th Int. conference on advances in web-based learning. 2005. P. 273-278.
33. A classifier based on a decision tree with verifying cuts / *J.G. Bazan, S. Bazan-Socha, S. Buregwa-Czuma, L. Dydo, W. Rzasa, A. Skowron* // The proceedings of the 23th Int. Workshop on Concurrency, Specification and Programming. 2014. P. 13-21.
34. *Aviad B., Roy G.* Classification by clustering decision tree-like classifier based on adjusted clusters // *Expert Systems with Applications*. 2011. Vol. 38. P. 8220-8228.

35. *Novak-Brzezinska A., Siminski R.* New inference algorithms based on rules partition // The proceedings of the 23th Int. Workshop on Concurrency, Specification and Programming. 2014. P. 164-175.
36. *Дьяконов А.Г.* Практикум на ЭВМ кафедры математических методов прогнозирования (логические игры, обучение по прецедентам): Учебное пособие. М.: Издательский отдел факультета ВМиК МГУ им. М.В. Ломоносова; МАКС Пресс, 2010. 164с.
37. *Дьяконов А.Г.* Практикум на ЭВМ кафедры математических методов прогнозирования (системы WEKA, RapidMiner и MatLab): Учебное пособие. М.: Издательский отдел факультета ВМиК МГУ им. М.В. Ломоносова; МАКС Пресс, 2010. 133с.
38. *Миниахметов Р.М., Цымблер М.Л.* Интеграция алгоритма кластеризации fuzzy c-means в PostgreSQL // Вычислительные методы и программирование. 2012. Т. 13. С. 46-52.
39. *Zhao Y. R* and Data Mining – Examples and Case Studies. Oxford: Elsevier Inc., Academic Press, 2012. 256 p.
40. *The elements of statistical learning. Data mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. Friedman.* NY: Springer, 2009. 745 p.

DIAGNOSTIC TESTS CLUSTERING IN ENGLISH PREPOSITIONS LEARNING ACCORDING TO BLOOM'S TAXONOMY

© 2015 S.A. Prokhorov, S.A. Suchkova, I.M. Kulikovskikh

Samara State Aerospace University named after Academician S.P. Korolyov
(National Research University)

The paper studies diagnostic test clustering in English preposition learning in terms of adaptive learning. The proposed clustering technique is based on Bloom's taxonomy that suggests creating diagnostic test banks according to cognition levels. We introduced a set of definitions and notations to present the clustering technique and "multiple choice" algorithms. We conducted a series of computational experiments using modeled answers to support the theoretical outcomes by k-means, c-means, and hierarchical clustering methods.

Key words: preposition, Bloom's taxonomy, adaptive learning, machine learning, clustering, cognition level, diagnostic test.

Sergej Prokhorov, Doctor of Technics, Professor, Head at the Information Systems and Technologies Department.

E-mail: sp.prokhorov@gmail.com

Svetlana Suchkova, Candidate of Philology, Associate Professor at the Foreign Languages Department.

E-mail: s.suchkova@gmail.com

Ilona Kulikovskikh, Candidate of Technics, Associate Professor at the Information System and Technologies Department.

E-mail: kulikovskikh.i@gmail.com