

DOI: 10.51790/2712-9942-2023-4-2-06

ФОРМИРОВАНИЕ СИНОНИМИЧНЫХ РЯДОВ С ЭКСПЕРТНОЙ ОЦЕНКОЙ ДЛЯ ПОЛУЧЕНИЯ КОЭФФИЦИЕНТОВ ЭМОЦИОНАЛЬНОСТИ СЛОВ

А. Р. Гончаров^а, С. А. Лысенкова^б, А. С. Назин^в

Сургутский государственный университет, г. Сургут, Российская Федерация

^а  mr.tesanta@ya.ru, ^б lsa1108@mail.ru, ^в nazin_as@surgu.ru

Аннотация: представлен результат экспертной оценки машинного перевода онлайн-сервисом «Reverso Context» в задаче формирования синонимичных рядов. Данные ряды формируют кортежи слов, предназначенные для проведения процедуры анкетирования методом «наилучшее-худшее масштабирование» и расчёта коэффициентов эмоциональности слов в новом тональном словаре. Номинальные оценки экспертов были переведены в связанные ранги, на основании которых рассчитан коэффициент согласованности экспертов. Обработанные оценки экспертов показывают адекватность подобранного данным сервисом перевода.

Ключевые слова: сентимент-анализ, словарь эмоций, сервис множественного перевода, экспертная оценка, машинный перевод, коэффициент согласованности.

Для цитирования: Гончаров А. Р., Лысенкова С. А., Назин А. С. Формирование синонимичных рядов с экспертной оценкой для получения коэффициентов эмоциональности слов. *Успехи кибернетики*. 2023;4(2):40–46. DOI: 10.51790/2712-9942-2023-4-2-06.

Поступила в редакцию: 18.06.2023.*В окончательном варианте:* 19.06.2023.

EXPERT EVALUATION-BASED GENERATION OF SYNONYMIC SETS FOR SENTIMENT ANALYSIS

A. R. Goncharov^a, S. A. Lysenkova^b, A. S. Nazin^c

Surgut State University, Surgut, Russian Federation

^a  mr.tesanta@ya.ru, ^b lsa1108@mail.ru, ^c nazin_as@surgu.ru

Abstract: this study presents the expert evaluation of machine translation by the Reverso Context online service as applied to the generation of synonymic sets. These sets are used for a “best-worst” survey and estimation of emotionality levels to make a sentiment dictionary. The expert scores were converted into ranks. We also estimated the consistency index for the experts. The expert evaluations indicated that the machine translation offered by the online service is adequate.

Keywords: comparative analysis, speech mood dictionary, machine translation, sentiment analysis, consistency index.

Cite this article: Goncharov A. R., Lysenkova S. A., Nazin A. S. Expert Evaluation-Based Generation of Synonymic Sets for Sentiment Analysis. *Russian Journal of Cybernetics*. 2023;4(2):40–46. DOI: 10.51790/2712-9942-2023-4-2-06.

Original article submitted: 18.06.2023.*Revision submitted:* 19.06.2023.

Введение

Мнения являются неотъемлемой частью человеческого общения. Под мнением подразумевают суждение о каком-либо объекте, в результате которого выражается эмоциональная окраска (тональность) автора и несущая смысловая идея текста [1]. Умение распознавать тональность текстов успешно применяется в потребительской сфере, в новостных ресурсах и в прочем социальном инструментарии. Тенденция значимости мнений обусловлена возникновением больших данных и онлайн-технологий, что подразумевает автоматизированное нахождение тональности (сентимент-анализ) и обработку электронных текстовых массивов.

Чаще всего тональность текста определяется классификацией — соотношением его по заданным категориям: ранжированная шкала или принадлежность к какому-либо классу эмоциональности в выделенном множестве («положительная», «нейтральная» или «отрицательная»).

Автоматизация позволяет ускорить процесс классификации, и одним из существующих методов автоматизированного сентимент-анализа является использование тональных словарей — заранее

составленных лексических шаблонов с их тональной оценкой, которая может быть выражена номинально в виде шкалы проявлений эмоций и/или числовым значением. Суть методики заключается в присвоении каждому слову текста тональной оценки при наличии данного элемента в словаре. Совокупность таких оценок используют в принятии решения об итоговой тональности источника.

Каждый элемент словаря относят к конкретной области применения, что подразумевает анализ определённых текстовых источников и сложность самостоятельного составления подобного словаря. Поэтому в информационных системах сентимент-анализа чаще всего используют готовые решения. Например, краудсорсинговый тональный словарь «NRC-EIL» [2], содержащий часто употребляемые англоязычные слова из социальной сети «Twitter» с коэффициентами интенсивности эмоционального тона. Словарь применяется в таких областях, как разведка мнений, оценка тенденций продвижения коммерческих Интернет-продуктов, анализ лозунгов политических партий. Также словарь используется в задачах генерации текстов и лирической поэзии [3].

Важно отметить, что формирование мнений и образов восприятия мира у человека происходит на сознательном уровне той культуры, в которой он воспитывался, рос и контактировал с другими её представителями. Язык общения содержит большой объём когнитивных моделей, описывающих «концептуальную вселенную» человека и передающих эмоциональный опыт и специфику культуры. Данные модели уникальны для каждой культуры, так как в основе мировоззрения каждого народа лежит своя система социальных ценностей и стереотипов. Поэтому результат сентимент-анализа на одном языке может не быть перенесён на другой язык.

Тональные словари в разных языках неодинаковы: разный состав и объём описательной лексики, хотя сами по себе передающие психологические состояния (эмоции) элементов универсальны. Несмотря на то, что исследуемый тональный словарь NRC-EIL был переведён на более, чем 100 других языков (включая русский язык), можно встретить примеры «уникальных» слов (рис. 1), используемых в определённых социальных группах.

	Слово	Эмоция	Интенсивность эмоции	Перевод
164	fumin	anger	0.779	фумин
193	stfu	anger	0.750	стфу
3824	frankenstorm	fear	0.672	Франкеншторм
6152	carol	joy	0.328	Кэрол
6416	chow	joy	0.135	чау
6417	pitter	joy	0.134	питтер
9528	fain	trust	0.398	файн

Рис. 1. Примеры «уникальных» слов в NRC-EIL. Примечание: составлено авторами по источнику [4]

В связи с этим была выдвинута гипотеза о межкультурном различии интерпретации слов NRC-EIL в русскоязычной культуре, результатом которого является изменение значений коэффициентов тональности в данном тональном словаре [5]. Для проверки существования выдвинутой теории необходимо провести анкетирование выделенной группы русскоязычной аудитории методом «наилучшее-худшее масштабирование» с применением синонимичных рядов, позволяющих респонденту уникально идентифицировать элементы и произвести выбор крайних элементов.

Материалы и методы

Главной особенностью проведения анкетирования методом BWS (best-worst scaling, «наилучшее-худшее масштабирование») [6] для расчёта коэффициентов эмоциональности слов является предоставление респонденту выбора крайних элементов (наилучший и наихудший) по рассматриваемому свойству из набора альтернатив. Метод опирается на частоту выбора объектов, которая обеспечивает метрику их сравнения важности, и в его основе лежит теория выбора по шкале максимального различия (maxdiff). Стоит отметить, что метод «BWS» и модель «maxdiff» не являются тождественными и рассматривают разные подходы в выборе: первый предполагает, что индивид производит

выбор наилучшего/наихудшего элемента в определённом порядке, в то время как второй предполагает одновременный выбор пары элементов на основе максимальной разности шкалы полезности [7].

Для подсчёта коэффициентов русскоязычной версии NRC-EIL участнику анкетирования поочередно предоставляют вариант кортежа из $n = 4$ разных слов из множества N и предлагают произвести крайний выбор двух элементов, где проявляется соответственно наилучшая и наихудшая интенсивность рассматриваемой эмоции (в данном случае исследуется категория «радость»). Для проведения исследования выбрана эмоциональная группа «радость» в связи с содержанием небольшого количества элементов и трудоёмким процессом получения коэффициентов. Многие эксперты сходятся во мнении, что «радость» входит в группу базовых эмоций человека [8].

Данная постановка задачи формирует полное ранжирование объектов в «лёгкой» для респондента форме, а исследователю предоставляются парные оценки. В связи с тем, что респондентам необходимо предоставить кортеж из элементов русскоязычного перевода словаря, возникает необходимость идентификации одинаковых элементов, так как в рассматриваемой эмоциональной группе было выявлено 139 дублирующих позиций (рис. 2).

	Слово	Эмоция	Интенсивность эмоции	Перевод
5286	love	joy	0.828	любовь
5302	greatful	joy	0.816	благодарный
5328	lovinlife	joy	0.797	любовь
5332	grateful	joy	0.789	благодарный
5418	thankful	joy	0.727	благодарный
5616	amour	joy	0.607	любовь

Рис. 2. Примеры дублирующих позиций в переводе NRC-EIL. Примечание: составлено авторами по источнику [4]

Некоторые исходные слова словаря, например, «grateful» и «thankful», имеют одно и то же значение в русскоязычном переводе («благодарный»). Данная лексическая особенность связана с разнообразием семантических установок исходного языка, которые не всегда имеют смысловые соответствия в системе содержания языков перевода. Стоит учесть, что не во всех случаях можно подобрать идеальный эквивалент слову оригинала.

Предоставление участнику анкетирования кортежа слов с одинаковыми элементами приведёт к неясности, нарушению концепции метода BWS и, в конечном итоге, к несогласованным итоговым результатам, в ходе исследования респондент не сможет определить смысловое различие предоставляемых элементов выбора.

Было принято решение обогатить перевод соответствующими эквивалентными значениями, которые призваны дополнить «центральный элемент мысли», образуя тем самым синонимичный ряд. Под «синонимичным рядом» подразумевается список однозначных слов, ранжированных в зависимости от интенсивности или оттенка определённого качества [9]. В данном исследовании синонимичный ряд отражает более или менее передающие понятия «опорного» слова (доминанта), расположенного для удобства в начале ряда. Например, для опорного слова «счастье» сформированный синонимичный ряд может иметь следующий вид:

< счастье, благодать, успех, блаженство >

Наполнение ряда является субъективным действием, не поддающимся строгой оценке правильности подбора соответствующих слов. Уровень знаний и личный образ восприятия мира у человека создают его личную интерпретацию вопроса при коммуникации, а образы ситуаций реагирования на окружение мотивируют использовать синонимические ряды в естественном языке.

Одним из способов составления синонимичного ряда является применение множественного

перевода с помощью аналогов, когда иностранному слову соответствует несколько значений с последующим выбором наиболее подходящих по контексту слов [10].

Правильный перевод и подбор соответствующих аналогов является трудной задачей для полной передачи смыслового значения исходного слова. Для получения адекватных результатов исследования привлекаются эксперты, оценивающие сформированный машинный перевод. Экспертам предоставляется пояснительная записка, отражающая цель и задачи экспертизы, приводятся краткие сведения о рассматриваемом предмете оценивания.

В оценке перевода могут быть задействованы эксперты с возможным наличием различающихся взглядов, концепций, опыта, поэтому возникает вероятность проблемы согласованности. В данном случае под «согласованностью» понимается общий уровень знаний экспертов, позволяющий прийти к единому заключению по решаемой задаче. Для расчёта согласованности экспертов применяется коэффициент согласованности (конкордации) Мориса Кендалла. Коэффициент выражает количественную степень близости индивидуальных мнений/ответов экспертов и, в данном случае, равен единице, если ответы экспертов одинаковы, и равен нулю, если ответы различны. Условие применения данного коэффициента заключается в представлении оценок каждого эксперта в виде рангового ранжирования: номинальные оценки каждого эксперта переводят в связанные ранги в порядке возрастания относительно других его значений. В случае одинаковых оценок присваивается ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны. Например, при оценивании экспертом некоторых объектов по пятибалльной шкале формируется ряд номинальных оценок $\langle 2, 1, 4, 5, 4 \rangle$ и при его переводе в ранжированные связанные ранги образуется ряд $\langle 2, 1, 3.5, 5, 3.5 \rangle$.

Итоговая согласованность экспертов по коэффициенту Кендалла с ранжированными связанными рангами рассчитывается по следующей формуле [11]

$$W = \frac{12 \cdot S}{m^2 (n^3 - n) - m \sum_{j=1}^m T_j}, \quad (1)$$

где: m – количество экспертов ($m = 3$); n – количество оценок ($n = 1264$); $S = \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} - \bar{x} \right)^2$; x_{ij} – оценка i -ого наблюдения j -ым экспертом; $\bar{x} = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} \right)$; $T_j = \sum_{k=1}^{H_k} \left(h_k^2 - h_k \right)$.

В формуле (1) T_j – показатель связанных рангов в j -ой ранжировке эксперта, H_k – число групп равных рангов в j -ой ранжировке, h_k – число равных рангов в k -ой группе связанных рангов при ранжировке j -ым экспертом. В данном случае $H_k = 4$ для каждой ранжировки ответов эксперта, так как в качестве оценивания использовалась номинальная шкала из 4 вариантов оценки перевода.

Как известно, одной из целей проведения экспертизы является получение полной согласованности экспертов. Необходимо убедиться, что полученный коэффициент согласованности W является не случайной величиной, а истинным значением, которому можно доверять в полученных результатах экспертизы. Поэтому произведём проверку значимости коэффициента согласованности. Для определения значимости связанных ранжированных рангов при большом числе наблюдений/оценок используется распределение хи-квадрат (χ^2):

$$\chi^2_{\text{расч}} = \frac{12 \cdot S}{m \cdot n \cdot (n + 1) - \frac{1}{n-1} \sum_{j=1}^m T_j}. \quad (2)$$

Полученное значение сравнивается с табличным $\chi^2_{\text{табл}}$ по заданному уровню доверительной вероятности p и числу степеней свободы $df = n - 1$. Обычно доверительная вероятность принимается равной 0.95. Для оценки значимости коэффициента конкордации W необходимо и достаточно, чтобы найденное значение было больше табличного ($\chi^2_{\text{расч}} > \chi^2_{\text{табл}}$). В таком случае гипотеза о согласии экспертов в ранжировках принимается, иначе имеется несогласованность экспертов.

Результаты и их обсуждение

Для формирования синонимичных рядов элементов группы «радость» словаря NRC-EIL был использован онлайн-сервис «Reverso Context» множественного перевода, который выдаёт список часто

используемых эквивалентных значений (рис. 3). Для удобства восприятия полученные ряды имеют глубину до 4 первых часто используемых элементов. Отмечено, что не все термины были распознаны сервисом, поэтому принято решение в таких случаях использовать в анкетировании только исходный русскоязычный перевод словаря.

	Слово	Синонимичный ряд
0	happiest	самый счастливый, самые счастливые, самым счас...
1	happiness	счастье, счастье, радость, благополучие
2	bliss	блаженство, счастье, радость, благодать
3	celebrating	празднование, праздник, посвященный, празднует
4	jubilant	ликующий, ликовали, ликовал, ликуют
5	ecstatic	восторженный, экстатический, экстатичный, экст...
6	elation	восторг, эйфория, радость, душевный подъём
7	beaming	сияющий, испускающий лучи, сияние, группировка
8	bestdayever	самый лучший день,
9	loveee	любимая,

Рис. 3. Примеры сформированных синонимичных рядов. Примечание: составлено авторами по источнику [4]

Для проверки адекватности подобранных онлайн-сервисом множественного перевода терминов, которые должны отражать смысловую идею первого исходного слова в ряду, были приглашены три эксперта.

В результате проведения экспертизы получены три ряда номинальных оценок перевода (рис. 4), значения которых располагаются в диапазоне от 0 до 3, где:

- «0» — полностью некорректный;
- «1» — скорее некорректный, чем корректный;
- «2» — скорее корректный, чем некорректный;
- «3» — полностью корректный.

	Слово	Машинный перевод	эксперт 1	эксперт 2	эксперт 3
0	happiest	самый счастливый, самые счастливые, самым счас...	3	3	3
1	happiness	счастье, счастье, радость, благополучие	3	3	3
2	bliss	блаженство, счастье, радость, благодать	2	3	3
3	celebrating	празднование, праздник, посвященный, празднует	2	3	2
4	jubilant	ликующий, ликовали, ликовал, ликуют	2	3	2
5	ecstatic	восторженный, экстатический, экстатичный, экст...	3	3	3
6	elation	восторг, эйфория, радость, душевный подъём	3	3	3
7	beaming	сияющий, испускающий лучи, сияние, группировка	2	3	2
8	bestdayever	самый лучший день,	3	2	3
9	loveee	любимая,	2	3	0

Рис. 4. Примеры полученных экспертных оценок. Примечание: составлено авторами по источнику [4]

Так как экспертные оценки измеряются в номинальной шкале, то их значения были переведены в ранжированные связанные ранги (рис. 5).

Была рассчитана согласованность экспертов, где коэффициент $W = 0,6697$.

	Слово	Машинный перевод	эксперт 1	эксперт 2	эксперт 3
0	happiest	самый счастливый, самые счастливые, самым счас...	729.5	735.0	708
1	happiness	счастье, счастье, радость, благополучие	729.5	735.0	708
2	bliss	блаженство, счастье, радость, благодать	123.5	735.0	708
3	celebrating	празднование, праздник, посвященный, празднует	123.5	735.0	105
4	jubilant	ликующий, ликовали, ликовал, ликуют	123.5	735.0	105
5	ecstatic	восторженный, экстатический, экстатичный, экст...	729.5	735.0	708
6	elation	восторг, эйфория, радость, душевный подъём	729.5	735.0	708
7	beaming	сияющий, испускающий лучи, сияние, группировка	123.5	735.0	105
8	bestdayever	самый лучший день,	729.5	124.5	708
9	loveee	любимая,	123.5	735.0	105

Рис. 5. Пример перевода оценок в связанные ранги. Примечание: составлено авторами по источнику [4]

По полученным оценкам была проведена проверка согласованности экспертов для убеждения в наличии общего уровня их знаний по решаемой задаче. При наличии связанных рангов $\chi^2_{\text{расч}} = 2537,51$. Табличное значение в данном случае $\chi^2_{\text{табл}} = 1346,79$ при $\alpha = 0,05$ и числе степеней свободы $df = 1263$. Поскольку $\chi^2_{\text{расч}} \geq \chi^2_{\text{табл}}$, то можно утверждать с 95% вероятностью, что экспертные оценки перевода взаимно непротиворечивы и имеется согласованность мнений экспертов по вопросу качества перевода.

Преобразуем оценки экспертов в два качественных класса перевода: «корректный» и «некорректный», где в «корректный» входят оценки $y \in \{2,3\}$, а в «некорректный» — $y \in \{0,1\}$. Количество элементов класса «корректный» составляет 1218 (при учёте голосования выбора из всех оценок экспертов), что составляет 96,36% от всего объёма словаря.

Заключение

Полученные оценки экспертов показывают адекватность подобранного перевода в сформированных синонимичных рядах из слов группы эмоции «радость» словаря NRC-EIL. Машинный перевод онлайн-сервисом «Reverso Context» показал способность подобрать эквивалентные значения по заданному опорному слову. Отмечено, что полностью с задачей формирования синонимичных рядов сервис не справился, что показывает необходимость ручного контроля для получения более достоверных результатов исследования.

Расчёт новых коэффициентов эмоциональности слов группы «радость» NRC-EIL основывается на попарных сравнительных оценках BWS соответствующих элементов словаря. Полученные синонимичные ряды позволят сформировать анкеты и провести опрос среди русскоязычных пользователей. Анкетирование по методу BWS позволит получить данные коэффициенты и частично проверить существование выдвинутой гипотезы о межкультурном различии в использовании данного тонального словаря.

Наличие словаря улучшает качество работы систем сентимент-анализа, снижает зависимость от внешних данных (признаков) и позволяет в удобной форме хранить оценочную лексику языка. Важно отметить, что полностью отказаться от использования внешних источников в глубоком сентимент-анализе невозможно, так как наличие дополнительных признаков снимает с текстов неоднозначность и уточняет смысл конструкций при помощи адаптации систем к рассматриваемой предметной области.

ЛИТЕРАТУРА

1. Юрганов А. А. Сентимент-анализ как инструмент исследования текстов. *Проблемы современной науки и образования*. 2017;29:39–41. EDN: ZDMSVF.
2. *The NRC Emotion Intensity Lexicon*. Режим доступа: <https://www.saifmohammad.com/WebPages/AffectIntensity.htm>.
3. Bena B., Kalita J. K. Introducing Aspects of Creativity in Automatic Poetry Generation. *ArXiv*. 2002;02511.

4. *Обработка и анализ данных с Pandas. Блог о разработке на Python.* Режим доступа: <https://egorovegor.ru/pandas-obrabotka-i-analiz-dannyh-v-python/>.
5. Гончаров А. Р., Лысенкова С. А. Анализ английского лексикона интенсивности эмоций и формирование алгоритма построения его русскоязычной версии. *Вестник кибернетики.* 2022;4:29–36. DOI 10.34822/1999-7604-2022-4-29-36.
6. Flynn T. N., Marley A. A. J. Best-Worst Scaling: Theory and Methods. *Handbook of Choice Modelling.* Edited by S. Hess, A. Daly. Edward Elgar Publishing; 2014. P. 178–201.
7. Marley A. A. J., Louviere J. J. Some Probabilistic Models of Best, Worst, and Best-Worst Choices. *Journal of Mathematical Psychology.* 2005;49(6):464–480. Режим доступа: <https://doi.org/10.1016/j.jmp.2005.05.003>.
8. Пенский О. Г. Роботы и человек. *Вестник Пермского университета. Серия: Математика. Механика. Информатика.* 2020;3:50–55. Режим доступа: <https://e.lanbook.com/journal/issue/322388>.
9. *Теория перевода: учеб. пособие* / сост. Ж. С. Пустовалова. Казань: КГК им. Жиганова; 2015. 82 с. Режим доступа: <https://e.lanbook.com/book/180910>.
10. *Основы технического перевода: учеб. пособие* / сост. Ю. В. Титова, Т. В. Капустина. Ульяновск: УлГТУ; 2016. 170 с. Режим доступа: <https://e.lanbook.com/book/165029>.
11. Лубенец Ю. В. Альтернативный коэффициент конкордации при наличии связанных рангов. *Вестник Воронежского государственного технического университета.* 2021;17(1):40–45. DOI 10.36622/VSTU.2021.17.1.005. EDN: XBBVII.