

УДК 004.822

МЕТОД ОЦЕНКИ КАЧЕСТВА СЕМАНТИЧЕСКОЙ СЕТИ ТЕКСТА**Кушеева М.Н., Гомбожапова Т.Н., Аюшеева Н.Н.***ФГБОУ ВО «Восточно-Сибирский государственный университет технологии и управления»,
Улан-Удэ, e-mail: anndonir@gmail.com*

В статье представлен метод оценки качества семантической сети, построенной автоматически на основе анализа научного текста. Предложенный метод оценки является результатом комбинирования метода оценки качества семантических сетей Ground Truth и метрики оценки качества машинного перевода Meteor. В качестве эталонной сети в исследовании применяются онтология предметной области и семантическая сеть, построенная вручную экспертом. Единицей рассмотрения в семантической сети принята униграмма, которая представляет собой элемент семантической сети вида «термин – отношение – термин». Точность, отклик и среднее гармоническое значение сравнения построенной семантической сети с эталонной сетью рассчитываются по формулам метрики Meteor. При расчете итогового показателя оценки качества учитывается штраф, вычисляемый на основе разницы фрагментов сравниваемых семантических сетей. Для проверки метода разработан программный прототип, на вход которого подаются эталонная сеть и анализируемая сеть, а на выходе получаем рассчитанную оценку качества семантической сети. Результаты проведенных вычислительных экспериментов в целом доказывают корректность разработанного метода и позволяют рекомендовать его для создания качественных семантических сетей. Также анализ результатов вычислительных экспериментов показал, что в качестве эталонной семантической сети следует брать сеть, построенную вручную, а онтологии предметных областей не совсем подходят для их использования в роли эталона.

Ключевые слова: обработка естественного языка, семантическая сеть, метод оценки качества, метрика Meteor, метод Ground Truth

SEMANTIC NETWORK OF TEXT QUALITY EVALUATION METHOD**Kusheeva M.N., Gombozhapova T.N., Ayusheeva N.N.***East Siberia State University of Technology and Management, Ulan-Ude, e-mail: anndonir@gmail.com*

The article presents a semantic network quality evaluation method. The semantic network built automatically in terms of an analysis of the scientific text. The proposed assessment method is the result of combining method Ground Truth and the machine translation quality evaluation metric METEOR. Domain ontology and semantic network built manually by an expert used as a reference network in the study. The unit of the semantic network is the unigram, which is an element of the semantic network on the type «term-relation-term». The precision, recall and harmonic mean of the comparison the constructed semantic network with the reference network calculated using the formulas of the Meteor metric. When calculating the final value of the quality evaluation the penalty is considered. The penalty calculate on the bases the fragments of the semantic network differential. To test the method, a software prototype has been developed, the input of which is the reference network and the analyzed network, and the output is a calculated semantic network quality evaluation. The results of the computational experiments in general prove the correctness of the developed method and allow us to recommend it for the creation of qualitative semantic networks. Also, the analysis of the computational experiments results showed that the semantic network built manually by an expert should take as a reference semantic network, and domain ontology are not quite suitable for their use as a standard.

Keywords: natural language processing, semantic network, a method for assessing the quality, metric METEOR, method Ground Truth

В настоящее время в задачах обработки естественного языка, решение которых основывалось на использовании семантических сетей, находят все большее применение другие подходы, например методы машинного обучения. Вместе с тем семантические сети остаются актуальными для решения многих задач, среди которых оценка методов обработки естественного языка, построение инвентаря значений слов, связывание языковых ресурсов, семантический поиск и др. [1]. Основной причиной отказа от использования семантических сетей является сложность при их построении, связанная с большой трудоемкостью их создания. Применение автоматизированных подходов сокращает этот процесс, но повышает требования к контролю качества построенных семантических сетей.

Наиболее известны следующие методы оценки качества семантических сетей [2]:

– Ground Truth; в основе метода лежит применение образцовой семантической сети;

– End-to-end; метод оценивает качество семантической сети по улучшению качества работы приложения, которое ее использует;

– «вручную»; метод заключается в том, что эксперты оценивают степень соответствия концептов заданной предметной области, а также корректность отношений между ними.

Перечисленные методы оценки качества семантических сетей базируются на различных показателях качества, которые дают оценочное значение, не совсем четкое и не совсем приемлемое для решения задачи оценки семантической сети текста,

а в случае метода «end-to-end» – практически неприменимое. Поэтому для решения данной задачи необходимы метрики оценки качества результатов естественно-языковой обработки, среди которых можно выделить следующие:

– BLEU (bilingual evaluation understudy); алгоритм разработан компанией IBM и оценивает качество перевода по шкале от 0 до 100 на основании сравнения машинного перевода с человеческим и поиска общих слов и фраз, является одной из самых простых в использовании метрик оценки машинного перевода [3];

– NIST; алгоритм разработан в американском Национальном институте стандартов и технологий (National Institute of Standards and Technology). Если для получения высокой оценки BLEU важнее правильный порядок слов, то NIST выше оценивает правильный выбор лексики. Фактически от метрики BLEU данная метрика отличается системой штрафов за неверный перевод [4];

– METEOR; разработчик Language Technologies Institute Carnegie Mellon University Pittsburgh. В этой технике за единицу оценки принимается не n -грамма, а слово. В данной метрике принимается во внимание вариация переводов в виде флексии, синонимов и изложения одинакового содержания с помощью других слов [5].

Данные метрики в целом основаны на n -граммном моделировании с использованием информационного поиска [3]. Изучение и анализ описаний метрик, алгоритмов их расчета, примеров применения позволяет сделать вывод о целесообразности выбора метрики METEOR для оценки качества семантических сетей в связи с ее высокой корреляцией с человеческой оценкой и достаточно низкой степенью адаптации к другой сфере применения.

Выбранная метрика в комбинации с методом Ground Truth позволяет предложить приемлемый метод оценки качества семантических сетей ЕЯ-текста.

В направлении исследований вопросов автоматического построения семантических сетей существует проблема, заключающаяся в следующем. Как правило, для оценки качества семантических сетей используется метод экспертной оценки. Однако организационные мероприятия по применению данного метода весьма трудоемки, так как формирование группы экспертов является многошаговым процессом, требующим проверки согласованности экспертов.

Решение этой проблемы возможно, если разработать автоматический метод оценки качества семантических сетей.

Исходя из этого, можно сделать вывод, что данная тема является достаточно актуальной, поскольку оценка качества семантических сетей значительно упростит задачу построения качественных семантических сетей, выведет компьютерную обработку текстов на новый уровень, а также совершит хоть и небольшой, но все-таки шаг навстречу новым разработкам в сфере искусственного интеллекта.

Таким образом, объектом исследования являются семантические сети, а предметом исследования – метод оценки их качества. Основной целью данного исследования является адекватная оценка правильности построения семантической сети.

Описание метода

Метод Ground Truth основан на использовании эталонной семантической сети, с которой сравнивается построенная сеть. Чаще всего в качестве эталонной сети принимаются тезаурусы, отличительным свойством которых является ограниченное множество типов отношений. В это множество не входят функциональные отношения, которые составляют большую часть выделяемых отношений в текстовом источнике. Поэтому тезаурусы для оценки качества семантической сети текста принимать за эталон нецелесообразно. За эталонную сеть можно принять либо полную онтологию предметной области, либо семантическую сеть, построенную вручную экспертом. Далее будем называть сетью-эталоном – эталонную сеть, сетью-кандидатом – сеть, построенную автоматически, которую необходимо сравнить с эталонной.

Метрика оценки качества машинного перевода METEOR за единицу оценки принимает не n -грамму, а слово или униграмму [3]. В семантической сети за такую униграмму будем принимать элемент семантической сети «термин – отношение – термин», поскольку именно он является наиболее примитивной фигурой при визуальном представлении семантической сети (рис. 1).

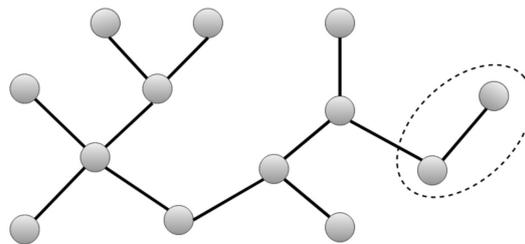


Рис. 1. Визуальное представление униграммы

Для определения схожести двух семантических сетей введем понятие – отображенные униграммы. Отображенные униграммы – это униграммы, которые присутствуют как в эталонной сети, так и в сети-кандидате. Количество отображенных униграмм является основной переменной в формулах расчета качества и играет важную роль в определении конечного значения схожести.

Точность, отклик и среднее гармоническое значение определяются по тем же формулам, которые используются метрикой METEOR для оценки качества машинного перевода. Точность (*P*-precision) униграмм семантических сетей есть отношение количества отображенных униграмм к общему числу униграмм в сети-кандидате (1).

$$P = \frac{m}{w_t}, \quad (1)$$

где m – количество отображенных униграмм; w_t – общее количество униграмм в сети-кандидате.

Данный коэффициент показывает, какую часть совпадающие униграммы составляют от общего числа униграмм в сравниваемой сети. Коэффициент отклика (*R*-recall) показывает отношение количества отображенных униграмм к общему числу униграмм сети-эталона, высчитывается по формуле

$$R = \frac{m}{w_r}, \quad (2)$$

где m – количество отображенных униграмм; w_r – общее количество униграмм в сети-эталоне.

Комбинируя точность и отклик, вычисляется среднее гармоническое значение по формуле

$$F_{mean} = \frac{10PR}{R + 9P}. \quad (3)$$

Метрика METEOR для оценки качества машинного перевода использует систему штрафов, начисляемых за несовпадающие фрагменты. В машинном переводе фрагментом называется часть предложения, которая совпадает у машинного (автоматического) и справочного (сделанного человеком) переводов. Чтобы учесть более длинные совпадения, метрика METEOR рассчитывает штраф, группируя униграммы в наименьшее возможное количество фрагментов, то есть чем длиннее n -граммы, тем меньше фрагментов, и в крайнем случае, когда вся строка машинного перевода соответствует справочному переводу, остается

только один фрагмент. В случае с семантическими сетями фрагментом будем считать такую часть семантической сети, в которой один термин имеет два и более отношений с другими различными терминами (рис. 2).

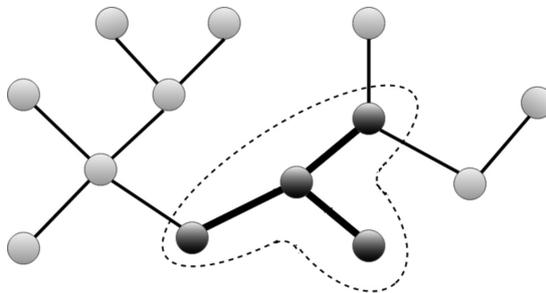


Рис. 2. Визуальное представление фрагмента

Таким образом, чем больше фрагментов имеет сеть, тем более информативной и полной она является. Поэтому штраф будет начисляться тогда, когда количество фрагментов сетей отлично друг от друга. Соответственно, чем меньше разница, тем меньше штраф, как и в исходной метрике. Таким образом, при вычислении фрагментации в формуле вычисления штрафа (*p*-penalty) в числителе будет разность количества фрагментов:

$$p = 0,5 \left(\frac{|c_1 - c_2|}{w_r + w_t} \right)^3, \quad (4)$$

где w_t – количество униграмм в кандидате; w_r – количество униграмм в эталоне; c_1 – количество фрагментов в эталонной сети; c_2 – количество фрагментов в сети-кандидате.

Еще одной отличительной чертой является то, что в знаменателе дроби, возводимой в куб при вычислении штрафа, будет находиться сумма общего количества униграмм обеих сетей, а не количество отображенных униграмм. Связано это с тем, что при построении сетей может отсутствовать достаточное число отношений, а значит, количество фрагментов может в разы отличаться, что приведет к вычислению слишком большого штрафа и впоследствии отрицательного итогового значения. Но количество фрагментов сети, а тем более и их разность, априори не может превышать сумму униграмм обеих сетей, а значит, итоговое значение и штраф будут принадлежать интервалу $[0,1]$ при любых возможных вариантах построения сети-кандидата. С математической же точки зрения,

фрагментация сети – это определение ее полноты. Отношение разности количества фрагментов к сумме общего числа униграмм, по сути дела, определяет, во сколько раз количество недостающих фрагментов могло бы увеличить полноту семантической сети.

Итоговая оценка качества семантической сети-кандидата рассчитывается по формуле

$$Score = F_{mean} (1 - p). \quad (5)$$

Данное значение представляет собой степень соответствия сети-кандидата эталонной семантической сети. Так как значение $Score \in [0,1]$ и по своей сути является оценкой силы корреляционной связи двух семантических сетей, для классификации полученного коэффициента наиболее верным решением будет использование шкалы Чеддока, представленной в табл. 1 [6].

Таблица 1

Шкала Чеддока

Значение $Score$	Сила связи
менее 0,3	слабая
от 0,3 до 0,5	умеренная
от 0,5 до 0,7	заметная
от 0,7 до 0,9	высокая
более 0,9	весьма высокая

Интерпретация итогового значения по шкале Чеддока позволит экспертам намного быстрее оценить степень совпадения сравниваемых семантических сетей и продолжить дальнейшие исследования.

Вычислительные эксперименты

Для проведения вычислительных экспериментов с целью определения корректности оценивания качества семантической сети разработанным методом был создан программный прототип. Рассмотрим результаты тестирования предложенного метода на примере научного текста не-

большого объема. Текст является узко специализированным, ориентированным на имеющуюся онтологию предметной области. После предобработки текст был загружен в программу построения семантической сети, в результате работы которой была получена сеть-кандидат. В качестве эталонной сети были взяты: построенная вручную сеть-эталон, достоверность которой была проверена экспертными методами оценки; фрагмент онтологии предметной области, которая была использована при построении семантической сети; фрагмент расширенной онтологии предметной области. При построении семантической сети использована онтология, сформированная на основе тезауруса RuТез. В расширенной онтологии присутствуют качественные отношения, такие как, например, функциональные.

В табл. 2 представлены результаты проведенных вычислительных экспериментов

В результате проведенных экспериментов можно сделать следующие выводы:

– наиболее подходящим эталоном для проверки правильности автоматического построения семантической сети является, как и было указано в описании метода, построенная вручную экспертом эталонная семантическая сеть;

– результат, полученный при сравнении сети-кандидата с используемой при построении сети онтологией, не может являться адекватной оценкой, поскольку штрафуются униграммы, которых нет в онтологии, а это униграммы с функциональными отношениями: чем их больше, тем больше будет и штраф, но значение штрафа не будет больше 10%. При рассматриваемом сравнении в расчет не берется ни правильное выявление подобных униграмм, потому что их не с чем сравнивать, ни то, что процент отображенных униграмм в таком случае будет очень высок, а значит, штраф максимально низок;

Таблица 2

Результаты вычислительных экспериментов

Сеть-эталон	Сеть-кандидат	Итоговая оценка $Score$	Сила связи	Время работы программы, с
Эталонная сеть, построенная вручную	Сеть-кандидат	0,6981	Заметная	0,003647
Расширенная онтология	Сеть-кандидат	0,2555	Слабая	0,004099
Онтология	Сеть-кандидат	0,5479	Заметная	0,003285
Расширенная онтология	Эталонная сеть, построенная вручную	0,3529	Умеренная	0,004287
Онтология	Эталонная сеть, построенная вручную	0,8434	Высокая	0,003822

– сравнение с расширенной онтологией устранило недостатки сравнения с использованной при построении сети онтологией. Однако в данном случае итоговая оценка будет очень низкой, поскольку число синонимов глаголов русского языка очень большое, а значит, вариаций униграмм с функциональными отношениями тоже. Но, по сравнению с принятием за эталон обычной онтологии, данную оценку можно считать достаточно релевантной, потому что здесь чем выше коэффициент, тем правильнее и информативнее построенная сеть;

– время выполнения алгоритма вычисления оценки качества семантической сети прямо пропорционально размеру оцениваемых сетей. Чем больше семантические сети содержат униграмм, тем, соответственно, больше время выполнения программы.

Заключение

Разработанный метод оценки качества семантической сети нуждается в тестировании на более серьезных текстах. Для его проведения необходим квалифицированный специалист и модифицированный программный прототип с расширенным функционалом. При выборе же тестовых данных необходимо учесть такие важные критерии, как стиль (текст должен быть научным или научно-публицистическим) и область знаний, которую охватывает текст (текст должен быть узкоспециализированным). Тестирование метода можно будет считать успешным

тогда, когда показатели работы программы будут идентичны расчетам оценки вручную экспертом. Однако результаты экспериментов доказывают корректность разработанного метода и позволяют рекомендовать его для создания качественных семантических сетей.

Список литературы

1. Усталов Д.А. Семантические сети и обработка естественного языка // Открытые системы. СУБД. 2017. № 2. С. 51–53.
2. Панченко А.И. Построение семантической сети из разнородных данных: дис. ... канд. тех. наук. Москва, 2010. 200 с.
3. Прохорова А.А. Анализ особенностей машинного перевода: Выпускная квалификационная работа (магистерская работа). СПб., 2016. 66 с.
4. Улиткин И.А. Автоматическая оценка качества машинного перевода научно-технического текста // Вестник МГОУ. Серия: Лингвистика. 2016. № 4. С. 174–182.
5. Иннеке де Корт. Оценка качества перевода с русского на нидерландский язык систем машинного перевода Google Translate и Bing Translator: Выпускная квалификационная работа (работа бакалавра). Гент. 2014. 62 с.
6. Марапов Д. Критерий корреляции Пирсона. [Электронный ресурс]. URL: <https://medstatistic.ru/theory/pirson.html> (дата обращения: 19.09.2019).
7. Шелманов А.О. Исследование методов автоматического анализа текстов и разработка интегрированной системы семантико-синтаксического анализа. М.: Изд-во ИСА РАН, 2015. 210 с.
8. Satanjeev Banerjee, Alon Lavie. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, June, 2005. P. 65–72. [Electronic resource]. URL: <https://www.cs.cmu.edu/~alavie/papers/BanerjeeLavie2005-final.pdf> (date of access: 19.09.2019).