

УДК 004.9

ИСПОЛЬЗОВАНИЕ НЕЧЁТКОЙ ЛОГИКИ ДЛЯ АНАЛИЗА ТЕКСТА

Поздняков М.В. («Национальный исследовательский университет ИТМО»)

Научный руководитель – кандидат физико-математических наук, доцент Иванов С.Е.
(«Национальный исследовательский университет ИТМО»)

Введение. С каждым годом увеличивается прирост новой доступной информации, в связи с чем требуется совершенствование существующих алгоритмов автоматической обработки текста. Среди различных используемых для этого математических аппаратов в том числе используется нечёткая логика. Наиболее часто модели на нечёткой логике решают задачи составления краткого изложения текста и определения эмоциональной составляющей.

Основная часть. Как правило, модели для обработки текста, основанные на аппарате нечёткой логики, можно разделить на следующие модули: предобработка текста, вычисление коэффициентов, фаззификация (при этом вычисленные ранее коэффициенты выступают в качестве входных данных), обработка на основании правил, дефаззификация, представление результатов.

Модели для краткого изложения текста как правило оперируют целыми предложениями и для шага фаззификации передают набор признаков предложений как входной набор. [2] Наиболее часто в их число входят наличие заглавных фразы в предложении, длина предложения, положение предложения в тексте, наличие тематических слов, сходство предложения с другими, вес термина, наличие имён собственных. [5] Дополнительно могут быть использованы среднее значение TF-ISF, сходство предложения с центроидом (вектор из усреднённых значений по тексту), наличие анафор и несущественной информации. [4]

Модели для определения эмоциональной составляющей оперируют отдельными словами. На вход на этапе фаззификации передаются как правило численный показатель слов с положительной и отрицательной окраской, показатель эмоциональных знаков препинания рядом со словом [3], привязка слова к эмоциональным категориям и связанные с этим значения [1].

Вне зависимости от целей, для которых применяются подобные модели, часто нечёткая логика является не единственным используемым математическим аппаратом. Так, почти 50% исследований было проведено с использованием гибридных моделей нечеткой логики с методами оптимизации. Также выполнялись исследования, посвященных гибридам нечеткой логики со статистическими и семантическими методами. [5] На этапе предобработки данных может выполняться кластеризация предложений или слов текста с применением растущей самоорганизующейся карты на основе графов [2], либо составление более простой для автоматической обработки иерархической структуры с использованием алгоритма агломеративной кластеризации [4]

В связи со спецификой выполняемых моделями задач, объективно оценить результат выполнения часто бывает затруднительно. В рассмотренных исследованиях для оценки результатов предлагается приглашать экспертов [3, 4], либо рассматривать полученные результаты в сравнении с другими методами (искусственные нейронные сети, векторные методы). [2]

Выводы. В заключении следует отметить, что подход нечёткой логики хорошо справляется с двусмысленностью и неточностью в произвольном тексте [1]. Именно обработка неопределенности во входных данных является наиболее сильной стороной данного аппарата по сравнению с другими моделями. Такой эффект достигается сопоставимостью работы аппарата с человеческими рассуждениями. [5] Гибкость нечёткой логики позволяет эффективно использовать её в сочетании с другими методами [2], а итоговые результаты получили высокие оценки в том числе от экспертов-лингвистов [4].

Список использованных источников:

1. Subasic, Pero & Huettner, Alison. (2001). Affect analysis of text using fuzzy semantic typing. *Fuzzy Systems, IEEE Transactions on*. 2. 483 - 496. 10.1109/91.940962.
2. Alfarrar, Mahmoud & Alfarrar, Abdalfattah & Alattar, Jamal. (2019). Graph-Based Fuzzy Logic for Extractive Text Summarization (GFLES). 96-101. 10.1109/ICPET.2019.00025.
3. Поздняков М.В., Осипов Н.А., Зудилова Т.В., Ананченко И.В., Иванов С.Е. Оценивание эмоциональной окраски текста при помощи нечёткой логики // *Современные наукоёмкие технологии -2022*. - № 10. - С. 32-36
4. F. Kyoomarsi, H. Khosravi, E. Eslami, P. K. Dehkordy and A. Tajoddin, "Optimizing Text Summarization Based on Fuzzy Logic," *Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008)*, Portland, OR, USA, 2008, pp. 347-352, doi: 10.1109/ICIS.2008.46.
5. A. Kumar and A. Sharma. Systematic literature review of fuzzy logic based text summarization // *Iranian Journal of Fuzzy Systems -2019 V.-16, №5*, pp. 45-59.