

Характеристика методов обработки неопределенности в лингвистике с использованием теории вероятностей и нечеткой логики

Characteristics of uncertainty processing methods in linguistics using probability theory and fuzzy logic

Кравец П.С.

Студент, Московский государственный технический университет им. Н. Баумана
e-mail: pkravets081@gmail.com

Kravets P.S.

Student, Bauman Moscow State Technical University
e-mail: pkravets081@gmail.com

Аннотация

Статья посвящена анализу применения теории вероятности и нечеткой логики для обработки неопределенности, часто возникающей при толковании значения смыслов слов. Подчеркивается, что комбинирование этих двух подходов даст наилучший результат для устранения лингвистической неопределенности.

Ключевые слова: теория вероятностей, нечеткая логика, неопределенность, лингвистическая неопределенность.

Abstract

The article is devoted to the analysis of the application of probability theory and fuzzy logic to the processing of uncertainty that often occurs when interpreting the meaning of words. It is emphasized that combining these two approaches will give the best result to eliminate linguistic uncertainty.

Keywords: probability theory, fuzzy logic, uncertainty, linguistic uncertainty.

Неопределенность формально определяется как недостаток знаний или недостаточная информация. Расплывчатость, двусмысленность, отсутствие конкретики или конфликт в получении информации затрудняют пользователю принятие решений или их исправление в условиях неопределенности.

Неопределенность прослеживается во всех областях и аспектах. Здесь мы говорим о существующей в лингвистике неопределенности, которая порождается утверждениями на естественном языке. Например, в русском языке одно слово может иметь разные значения и представлять разные части речи контекстуально в разных местах. Слово «пропасть» имеет разные значения в предложениях «Он хотел преодолеть эту пропасть» и «Эта вещь могла пропасть лишь по одной причине». В приведенных примерах «пропасть» – существительное и глагол одновременно. Такие двусмысленности обычно устраняются знанием контекста или здравым смыслом.

Работа с неопределенностью означает попытку получить наилучшее возможное решение для данной ситуации без какой-либо путаницы. Хотя люди легко справляются с этим, для машин это не так просто. Сарказм, разговорная речь, чрезмерное использование негативных слов и иногда не могут быть правильно поняты, что приводит к неверному определению того или иного слова.

Существует множество способов обработки неопределенности – рассуждения, нечёткая логика, фактор определенности, скрытая марковская модель, нейронные сети и т.д. В статье

рассматриваются два наиболее широко используемых метода – теория вероятностей и нечеткая логика.

Что касается существующих подходов, то в 2012 г. Л. Датт и М. Куриан изучали необходимость управления неопределенностью с использованием различных методов. Авторы дают классификацию различных типов неопределенности, преобладающих в реальном мире, и различных методов, таких как вероятностный анализ, нечеткий анализ, байесовский сетевой анализ и т.д., для обработки неопределенности. В статье также предлагается использовать вероятностный анализ и нечеткий анализ как наилучшие способы уменьшения лингвистической неопределенности [1].

А. Мегдади и др. предлагали использовать нечеткую логику для моделирования неопределенности [2]. Они также привели сравнение теории вероятностей и нечеткой логики, которые лучше всего подходят для работы с неопределенностью. Авторы пишут, что нечеткая логика используется для нестатистической неопределенности, а теория вероятностей – для статистической. Они также формулируют, что комбинация этих теорий может быть использована для обработки неопределенностей, которые одновременно существуют в системах реального мира. Авторы представляют математическую структуру под названием *модель дискретной статистической неопределенности (DSU)* для представления вероятностных нечетких систем.

А. Оже и Д. Рой, в свою очередь, учитывают меру неопределенности в текстовых данных [3]. В их статье описана двусмысленность в естественном языке, которая может быть разрешена с помощью лингвистических контекстов. Так авторы решают лингвистическую двусмысленность с помощью лексикона, грамматик, словаря и алгоритмов, отмечая, что используемый метод автоматической маркировки выражений объясняет, как можно получить доступ к определенности.

Л. Заде в своей работе излагает базовый подход к кванторам на естественном языке таким образом, что его можно интерпретировать как нечеткие числа [4]. Автор также объясняет применение семантики тестовых оценок к переводу нечетко выраженных возможностей, обеспечивающей метод вывода из баз знаний, которые содержат такие предположения, как в большинстве баз знаний реального мира. В его работе представлен метод уменьшения неопределенности в базах знаний естественного языка.

Ф. Грун и А. Мослех объясняют байесовские выводы как процесс, в котором представлению состояния присваивается возможное значение *true* или *false* на основе наблюдений [5]. Авторы также признают, что теорема Байеса не способна справиться с неопределенностью, поскольку само неопределенное наблюдение не позволяет вычислить вероятность наблюдения. В своей работе они преодолели это ограничение, переопределив функцию правдоподобия вероятности на вероятность того, что не противоречит наблюдению. Теория, предложенная в этой статье, применима к ситуациям, когда неопределенность не зависит от предварительных знаний об окружающей среде.

Ранее теорема нечетких множеств использовалась для решения проблемы лингвистической неопределенности, как упоминал Лотфи Заде [4], но мы будем использовать байесовский (база знаний) взгляд на теорию вероятности для решения лингвистической неопределенности.

Что касается лингвистической неопределенности, то она относится к неопределенности в языке. Неточная информация часто является неотъемлемой характеристикой человеческого мышления. Это возникает из высказываний, которые произносятся, или посредством прямого общения. Лингвистическая неопределенность бывает пяти основных типов [6]: *двусмысленность, неопределенность, зависимость от контекста, отсутствие конкретики и неопределенность теоретических терминов.*

Двусмысленность возникает из-за отсутствия релевантной информации, которая могла бы помочь устранить неопределенность. Двусмысленная неопределенность в основном возникает из-за концепции, что слово может иметь более одного значения: к примеру, фраза

«нужно купить гвоздики» порождает двусмысленность, поскольку неясно, идет ли речь о гвоздях или о цветах.

Все, что не является точным, считается расплывчатым [7]. Например, в контексте продуктов питания фраза «Я хочу съесть что-нибудь вкусненькое» не указывает чётко, чего он/она хочет. «Вкусненькое» может означать разные вещи для разных людей в зависимости от их предпочтений. Подобная *неопределенность* может быть устранена путем придания ей операционного значения или определения расплывчатого термина.

Контекстная зависимость относится к путанице между различными значениями, которые слово может принимать в зависимости от контекста. Так, слово «спутник» будет менять свое значение: «Мои спутники оказались милыми собеседниками», «Каменная соль – спутник нефти» или «Луна – спутник Земли». Сарказмы и каламбуры подпадают под эту категорию.

Отсутствие конкретики относится к недостаточному уточнению того, что именно подразумевается в той или иной фразе («я уеду далеко» – слово «далеко» с отсутствием конкретики порождает неопределенность), и *неопределенность теоретических терминов* – это множественное значение одних и тех же терминов в различных науках (термин «сущность» будет иметь разное значение, например, в физике и философии).

Теория вероятности существует с XVI в., когда Джероламо Кардано предпринял попытку проанализировать азартные игры. Вероятность в простых терминах – это вероятность/степень убежденности в том, что событие произойдет. *Теория вероятностей* – это средство анализа случайных событий.

Одним из основных расширений теории вероятностей является логика Байеса. Это производное из условной вероятности, которая определяет возможность наступления одного события при условии, что произошло другое событие.

В теории вероятностей логика Байеса определяет вероятность события на основе уже имеющегося знания связанных с ним условий. База знаний постоянно создается путем наблюдения, мониторинга и оценки предыдущего эксперимента или существующих фактов об эксперименте.

Основное значение теоремы Байеса заключается в том, что она основана на теоретическом доказательстве и использует условную вероятность для классификации событий. Более того, это один из немногих методов, способных к прямой оценке неопределенности, когда доступно мало данных или статистических выводов, если они есть [1].

В отличие от теории вероятностей, где результат является атомарным, т.е. истинным или ложным, нечёткая логика имеет дело со «степенью истинности» или «степенью принадлежности». Она присваивает любое действительное число между [0, 1], которое показывает, «насколько верно» условие.

Основатель нечёткой логики Лотфи Заде заметил, что человеческое мышление включает в себя целый ряд возможностей между четкими «Да» и «Нет», например, «возможно, нет», «не могу сказать», «безусловно, да» и т.д. [8]. Нечёткая логика напоминает человеческие рассуждения в том смысле, что она присваивает определенную степень истинности, включающую промежуточные возможности.

При создании нечёткой системы нужно чётко определить все функции принадлежности: здесь экспертные знания кодируются в нечёткие правила [5]. Система нечёткой логики хорошо справляется с расплывчатостью и неопределенностью, присущими человеческой базе знаний, и в результате позволяет нам представлять лингвистические термины.

В общем, теория вероятностей и нечёткая логика имеют разные области применения: несмотря на тесную взаимосвязь, ключевое различие заключается в том, для чего они лучше подходят.

Вероятность имеет дело с возникновением или ненаступлением событий, а не с фактами. Теория вероятностей не имеет ничего общего с рассуждениями. Это либо полностью верно,

либо ложно (атомарно). С другой стороны, нечёткая логика полностью зависит от степени истинности и отражает концепцию неопределенности.

Хотя байесовская сеть является одним из наиболее эффективных подходов к определению вероятности, у него есть несколько недостатков. Неизвестные значения, противоречивые знания и вычислительные затраты являются основными из них.

Главный недостаток нечёткой логики заключается в том, что она полностью зависит от контекста, а правила общего назначения довольно трудно получить. Это также требует, чтобы все функции принадлежности были чётко определены перед экспериментом.

Классическая теория вероятностей была успешной в механистических системах, где взаимодействия и зависимости чётко определены, таких как квантовая механика, системы связи, статистическая механика и смежные области. Но, когда это происходит в областях, где должны быть задействованы человеческие эмоции, восприятие и рассуждения, это терпит неудачу. Наилучший возможный способ справиться с подобным описан у Л. Заде, где говорится, что можно использовать как теорию вероятностей, так и нечёткую логику. Данный подход помогает обобщить классические законы теории вероятности и расширить сферу её применения.

Простыми словами, вероятностная логика определяет вероятность того, что событие окажется истинным, в то время как нечёткая логика определяет, насколько оно истинно. Как говорит Заде [5], эти два метода скорее дополняют друг друга, чем конкурируют.

Литература

1. Dutt L. S., Kurian M. Handling of Uncertainty A Survey // International Journal of Scientific and Research Publications, 2013. – Volume 3. – Issue 1. – P. 1-4.
2. Meghdadi A.H., Akbarzadeh T. M. R. Uncertainty modeling through probabilistic fuzzy systems //Fourth International Symposium on Uncertainty Modeling and Analysis, 2003. ISUMA 2003. – IEEE, 2003. – P. 56-61.
3. Auger A., Roy J. Expression of uncertainty in linguistic data //2008 11th International Conference on Information Fusion. – IEEE, 2008. – P. 1-8.
4. Zadeh L.A. A computational approach to fuzzy quantifiers in natural languages //Computational linguistics. – Pergamon, 1983. – P. 149-184.
5. Groen F.J., Mosleh A. Foundations of probabilistic inference with uncertain evidence //International Journal of Approximate Reasoning. – 2005. – Т. 39. – №. 1. – P. 49-83.
6. How Bayesian Inference Works, Vincent Granville. TechTarget. URL: <https://www.datasciencecentral.com/how-bayesian-inference-works/> (Дата посещения: 08.03.2023).
7. An Intuitive (and Short) Explanation of Bayes' Theorem, Better Explained. URL: <https://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/> (Дата посещения: 08.03.2023).
8. Beer M. A summary on fuzzy probability theory //2010 IEEE International Conference on Granular Computing. – IEEE, 2010. – P. 5-6.
9. Qin B. et al. A rule-based classification algorithm for uncertain data //2009 IEEE 25th international conference on data engineering. – IEEE, 2009. – 3. 1633-1640.