МЕТРИКА, ТОПОЛОГИЯ И ПЕРСИСТЕНТНОСТЬ В СМЫСЛОВОМ ПРОСТРАНСТВЕ

Ивашенко В. П.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: ivashenko@bsuir.by

Рассматриваются требования и предлагается подход к моделированию метрических свойств смыслового пространства, учитывающий топологические свойства, требования устойчивости и основанный на использовании вложения метрических пространств с ограничением максимально допустимого расстояния и их погружения в нормированные пространства. Рассмотрены свойства размерности таких пространств.

Введение

В ряде работ рассмотрены модели семантического [1] или смыслового пространства [2].

Такие пространства состоят из элементов, которые связаны метрически или топологически. Пространства могут в себя включать подпространства. Порядок включения подпространств и свойства пространства взаимно определяемы. В работе [2] рассмотрена метамодель пространства, которая позволяет определить порядок включения подпространств смыслового пространства, в соответствии с процессом становления онтологических структур, сводящемуся к добавлению в них новых элементов. Важным свойством является устойчивость метрических и топологических свойств при переходе от подпространства к надпространству.

I. Метрика смыслового пространства

Ранее рассмотрены топологические замыкания и соответствующие замкнутые множества обозначений в онтологиях, которые устойчивы (персистентны) по отношению к добавлению новых элементов онтологических структур, сводящихся к графовым струкурам [2].

Если топологические замыкания на графовых структурах тесно связаны с транзитивными бинарными отношениями (графовыми), то в отличие от квазиметрик (дистанций на орграфах), которые также связаны с транзитивными бинарными отношениями и (квази-равномерными) топологическими пространствами, метрики связаны с симметричными бинарными отношениями [3].

В силу того, что в языках, ориентированных на явное представление ассоциаций и метазаний, дистанция между знаками (обозначениями) всегда может быть сокращена добавлением элементов в онтологическую структуру до (конечного) расстояния, заданного некоторой ассоциативной связью, то можно сказать, что пространство «компактифицируется» до заданного «максимально допустимого» расстояния, с другой стороны в этом пространстве должны помещаться все знаки, входящие в любую онтологическую структуру, при этом расстояние между различными знаками должно быть ненулевым.

Одним из способов обеспечить эти требования, не использующим нецелочисленные значения для расстояний и координат, является рассмотрение такого пространства как многомерного, в котором все знаки помещены в многомерный шар и метрика в котором задаётся как функция (максимум) метрик вдоль каждой из координат.

Если пространство нормировано, то можно ввести базис и векторное представление для такого метрического пространства. При этом видится целесообразным, чтобы базис определялся в рамках метрической модели однозначно (канонически), отталкиваясь от канонической последовательности канонических представлений устойчивых топологических пространств, что, однако, оставляет вопрос о метрических свойствах в таком пространстве и устойчивости его метрики по отношению к отображению на расширенное пространство, т.е. возможности вложения в надпространство. Этот вопрос легко решаем, возможно, только для метрики простейшего дискретного пространства, однако размерность такого пространства будет в общем случае избыточна. В ином случае проблема по всей видимости в том, что даже такая метрика может мало чем отличаться от простейшей метрики дискретного топологического пространства, не в плане отличий вообще, а в плане обладания устойчивыми отличительными свойствами.

Так как может быть произвольное количество знаков смыслового пространства, находящихся попарно на максимальном расстоянии, то в метрическом пространстве каждый из них является центром шара в некоторой метрике, на поверхности которого лежат остальные. Для метрики, порождённой нормой, размерность соответствующего нормированного пространства будет равна количеству таких знаков, уменьшенному на 1.

Утв. 1. Если использовать только целочисленные значения координат, то размерность пространства будет расти в худшем случае не менее чем линейно от количества знаков.

Минимальная размерность пространства по отношению к количеству вершин будет достигаться в случае гиперкубического графа.

Видится целесообразным, чтобы координатно-векторное представление знаков в смысловом пространстве было достаточным для восстановления дистанций между ними, каким оно было бы до «компактификации», т.е. для восстановления локальных квази-метрических свойств в топологически устойчивых подструктурах смыслового пространства, наподобие того, как представление в пространстве с («истинной») эвклидовой размерностью графа [4] достаточно для его восстановления с точностью до изоморфизма по набору точек в отличие от его представления в пространстве с (эвклидовой) размерностью графа [5]. В частности, чтобы не связанные инцидентностью в экстенсионале знаки не находились на том же расстоянии как и связанные.

Чтобы выполнить последнее условие о достаточности, сформулируем следующие требования: знаки, минимальные окрестности которых имеют непересекающиеся замыкания, должны лежать друг от друга на максимально допустимом расстоянии, размерность пространства при этом должна быть минимальна; расстояние между знаками, одно из замыканий минимальных окрестностей которых является подмножеством другого, либо равно максимальному допустимому расстоянию, либо на единицу больше количества замыканий между ними, если это количество меньше максимально допустимого расстояния.

В случае выражения локальных метрических свойств метрикой городских кварталов и с учётом остальных требований размерность пространства будет не меньше максимального расстояния (цепи), уменьшенного на 1, т.е. будет зависеть линейно.

Таким образом, в худшем случае затраты на хранение возрастут с полилоглинейных до квадратичных относительно количества знаков, чтобы это компенсировать возможно использования кратких структур данных [6].

II. Размерность смыслового пространства

Пусть m – максимально допустимое расстояние, а n – количество (различимых) знаков, тогда размерность d пространства без учёта условия о достаточности:

$$d = \max \quad \left(\{1\} \cup \left\{ \left\lfloor \frac{(n-2)}{2*(m-1)} \right\rfloor \right\}\right)$$

Количество n можно выразить:

$$n \leq (1 + \mathrm{sign}(d-1)) * (m-1) * d + 2$$

Размерность D пространства с учётом условия о достаточности:

$$D \leq \max \quad (\{1\} \cup \{n-m\})$$

Заключение

Сформулированы требования и предложен подход к моделированию метрических свойств смыслового пространства, выявлены свойства размерности смыслового пространства с учётом сформулированных требований.

В перспективе дальнейшему исследованию подлежат: вопрос о возможности повышения эффективности поиска за счёт моделирования метрических свойств смыслового пространства предлагаемым образом и вопрос о преимуществах или недостатках методов поиска за счёт моделирования метрических свойств по отношению к методам поиска за счёт моделирования топологических свойств смыслового пространства.

В случае онтологий, топология, свойства топологического пространства, связаны с отношением частичного порядка, которое соответствует модели полной или конфлюентной персистентности в задачах семантического протоколирования процессов (в т.ч. обработки знаний), для которых были разработаны соответствующие модели протоколирования и структуры [7].

Поэтому некоторые вопросы топологического плана могут решаться на основе предложенных соответствующих моделей и методов.

1. Manin, Yu.I., Marcolli, M. Semantic spaces. Math. Comput. Sci. – 2016. – 10, no.4, – 459–477 pp.
2. Ivashenko, V. Semantic space integration of logical knowledge representation and knowledge processing models / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS); ed.: V. V. Golenkov [et al.]. – Minsk, 2023. – Iss. 7. – C. 95–114.
3. Smyth, M., Quasi uniformities: reconciling domains with metric spaces / M. Main, A. Melton, M. Mislove, D. Schmidt (eds.) // Mathematical Foundations of Programming Language Semantics, Lecture Notes in Computer Science. – Springer-Verlag, 1988. – 298. – 236–253 pp.
4. Erdős P., Simonovits M. On the chromatic number of geometric graphs // Ars Comb. – 1980. – 9. – 229–246 pp.
5. Harary, F., Melter, R.A. On the metric dimension of a graph. Ars Combinatoria, 1976. – 2. – 191–195 pp.
6. Jacobson, G. Space-efficient static trees and graphs // 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. – Research Triangle Park, NC, USA, 1989. – 549-554 pp.
7. Ivashenko, V. Semantic Logging of Repeating Events in a Forward Branching Time Model / V. Ivashenko, N. Zotov, M. Orlov // Pattern Recognition and Information Processing (PRIP'2021) = Распознавание образов и обработка информации (2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus / United Institute of Informatics Problems of the National Academy of Sciences of Belarus. – Minsk, 2021. – 149–152 pp.