ПРИМЕНЕНИЕ АВТОКОДИРОВЩИКОВ И МЕТОДА АДАПТИРУЕМОГО КОЭФФИЦИЕНТА ОБУЧЕНИЯ ДЛЯ ПОНИЖЕНИЯ РАЗМЕРНОСТИ В СМЫСЛОВОМ ПРОСТРАНСТВЕ

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

Рассматривается применение нейросетевых моделей и методов понижения размерности данных, использующих адаптивный (адаптируемый) коэффициент обучения, об элементах многомерного метрического смыслового пространтва. Используется метод точного аналитического вычисления адаптивного коэффициента обучения для линейного автокодировщика.

Введение

Многомерные метрические смысловые (под)пространства [1] (размерности δ) на некотором X образуются при построении из множества псевдометрических смысловых пространств, в соответствии со следующими теоремами 1 и 2 [2], а также при переходе от квазиметрических признаков к метрическим в условиях ограничений на связность и структуру метасвязей (максимально допустимым расстоянием m) представленных в смысловом пространстве знаний, в соответствии с выражением [3]:

$$\delta = \max\left(\{1\} \cup \left\{\left\lfloor \frac{(|X|-2)}{2*(m-1)} \right\rfloor\right\}\right).$$

Теорема 1. Пусть на множестве X заданы псевдометрическое с псевдометрикой π и псевдометрическое с метрикой μ пространства, а также $\lambda \geq 1$. Тогда $\langle X,\rho\rangle$ – псевдометрическое пространство, где

$$\rho\left(\langle\chi,\gamma\rangle\right) = \left(\left(\mu(\langle\chi,\gamma\rangle)^{\lambda}\right) + \left(\pi(\langle\chi,\gamma\rangle)^{\lambda}\right)\right)^{\left(\lambda^{-1}\right)}$$

или

$$\rho\left(\langle\chi,\gamma\rangle\right) = \max\left(\{\mu\left(\langle\chi,\gamma\rangle\right)\} \cup \{\pi\left(\langle\chi,\gamma\rangle\right)\}\right).$$

Теорема 2. Пусть на множестве X заданы псевдометрические с псевдометрикой π и псевдометрикой μ пространства, а также $\lambda \geq 1$ и $(\neg(\chi = \gamma)) \rightarrow (\mu(\langle\chi,\gamma\rangle) + \pi(\langle\chi,\gamma\rangle) \geq 0)$. Тогда $\langle X,\rho\rangle$ – метрическое пространство, где

$$\rho\left(\langle\chi,\gamma\rangle\right) = \left(\left(\mu(\langle\chi,\gamma\rangle)^{\lambda}\right) + \left(\pi(\langle\chi,\gamma\rangle)^{\lambda}\right)\right)^{\left(\lambda^{-1}\right)}$$

или

$$\rho\left(\langle\chi,\gamma\rangle\right) = \max\left(\{\mu\left(\langle\chi,\gamma\rangle\right)\} \cup \{\pi\left(\langle\chi,\gamma\rangle\right)\}\right).$$

I. Задача понижения размерности пространства

Необходимость в понижении размерности многомерного смыслвого пространства связана с необходимостью применения производительных, эффективных, высокопродуктивных алгоритмов поиска элементов и структур в смысловом пространстве. Размерность пространства сравнимая с количеством элементов в нём (линейнозависящая от их количества) является достаточным условиям неэффективности подобных алгоритмов, поэтому требуется сократить размерность соответствующих данных.

Кроме этого следует учесть, что механизмы, включая модели и методы понижения размерности смыслового пространства, с целью интерпретируемости и объяснимости получаемых с помощью них результатов, сами должны быть частью смыслового пространства [1,2,3,4].

Дано: множество образов $X = \left\{X^{(i)} \middle| X^{(i)} \in V^n\right\}$ из V^n, $p \leq n$.

Требуется: найти такие e из $(V^p)^{(V^n)}$ и d из $(V^n)^{(V^p)}$, что $\sum_i \left(e \circ d\left(X^{(i)}\right) - X^{(i)}\right)^2 \rightarrow \min$.

Задача понижения размерности пространства связана с таким направлением, как многомерное шкалирование [5]. Одним из методов для решения такой задачи являются искусственные нейросетевые модели – автокодировщики [6], которые бывают разных видов – принимающие значения из множества элементов поля с положительной натуральной характеристикой и характеристикой 0 (действительнозначные, комплекснозначные и др.)

Следует отметить, что если задана метрика μ на X ($\mu \in D^{\left((V^n)^2\right)}$), то на $H = \left\{H^{(i)} \middle| \left(H^{(i)} = e\left(X^{(i)}\right)\right) \wedge \left(X^{(i)} \in V^n\right)\right\}$ также может быть задана метрика ρ:

$$\rho = \left\{\langle\langle\alpha,\beta\rangle, \langle d(\alpha), d(\beta)\rangle\rangle \middle| \langle\alpha,\beta\rangle \in (V^p)^2\right\} \circ \mu.$$

С целью представления их в смысловом пространстве среди комплекснознаных автокодировщиков будем выделять (комплекснозначные) рациональные автокодировщики. Все функции таких кодировщиков являются рациональными, включая функции активации f: линейная, (полу)ограниченная линейная, недоограниченная линейная (Leaky ReLU), рациональная сигмоида и другие:

$$f(x) = \frac{1+|x|+x}{1+|x|} * \frac{x}{2}.$$

В случае линейного автокодировщика e и d имеют вид:

$$\begin{array}{l} e(X^{(i)}) = X^{(i)} * W^{(1)} - T^{(1)} \\ d(H^{(i)}) = H^{(i)} * W^{(2)} - T^{(2)} \end{array}.$$

II. Обучение линейного автокодировщика с адаптивным коэффициентом обучения

Решение задачи понижения размерности пространства сводится к обучению автокодировщика $D \subseteq \mathbb{R}$.

Пусть:

$$\begin{array}{l}
H^{(i)} = \mathrm{X}^{(\mathrm{i})} * \mathrm{W}^{(1)} - \mathrm{T}^{(1)} \\
Y^{(i)} = H^{(i)} * \mathrm{W}^{(2)} - \mathrm{T}^{(2)} \\
D^{(i)} = X^{(i)} \\
\Delta\mathrm{T}^{(2)} = -\sum_i \left(D^{(i)} - Y^{(i)}\right) \\
\Delta\mathrm{T}^{(1)} = \Delta\mathrm{T}^{(2)} * \left(\mathrm{W}^{(2)}\right)^T \\
\Delta\mathrm{W}^{(2)} = \sum_i \left(H^{(i)}\right)^T * \left(D^{(i)} - Y^{(i)}\right) \\
\Delta\mathrm{W}^{(1)} = \left(\sum_i \left(X^{(i)}\right)^T * \left(D^{(i)} - Y^{(i)}\right)\right) * \left(\mathrm{W}^{(2)}\right)^T \\
\Delta H^{(i)} = \mathrm{X}^{(\mathrm{i})} * \Delta\mathrm{W}^{(1)} - \Delta\mathrm{T}^{(1)} \\
\Delta G^{(i)} = \Delta H^{(i)} * \Delta\mathrm{W}^{(2)} \\
\Delta Y^{(i)} = \Delta H^{(i)} * \mathrm{W}^{(2)} + H^{(i)} * \Delta\mathrm{W}^{(2)} - \Delta\mathrm{T}^{(2)} \\
e = \sum_i \left(Y^{(i)} - D^{(i)}\right)^2 \\
d = \sum_i \left(Y^{(i)} - D^{(i)}\right) \cdot \Delta Y^{(i)} \\
c = \sum_i \left(\Delta Y^{(i)}\right)^2 + 2 * \sum\limits_i \left(Y^{(i)} - D^{(i)}\right) \cdot \Delta G^{(i)} \\
b = \sum_i \left(\Delta Y^{(i)} \cdot \Delta G^{(i)}\right) \\
a = \sum_i \left(\Delta G^{(i)}\right)^2 \\
r = \frac{2*a*c-3*b^2}{12*a^2} \\
q = \frac{b^3-a*b*c+2*a^2*d}{8*a^3}
\end{array}$$

тогда адаптивный коэффициент обучения α:

$$\alpha \in \left\{ s - \frac{p}{s} - \frac{b}{2*a} \middle| s = \sqrt[3]{-q \pm \sqrt{r^3+q^2}} \right\}$$

и

$$\begin{array}{l} T^{(1)} \leftarrow T^{(1)} + \alpha * \Delta T^{(1)} \\ T^{(2)} \leftarrow T^{(2)} + \alpha * \Delta T^{(2)} \\ \mathrm{W}^{(1)} \leftarrow \mathrm{W}^{(1)} + \alpha * \Delta\mathrm{W}^{(1)} \\ \mathrm{W}^{(2)} \leftarrow \mathrm{W}^{(2)} + \alpha * \Delta\mathrm{W}^{(2)} \end{array}.$$

При этом область допустимых значений для α:

$$6 * a * \alpha^2 + 6 * b * \alpha + c \geq 0$$

, а ошибка достигает значения:

$$a * \alpha^4 + 2 * b * \alpha^3 + c * \alpha^2 + 2 * d * \alpha + e$$

.

В случае $a = 0$ α выражается из:

$$3 * b * \alpha^2 + c * \alpha + d = 0.$$

С целью обеспечения объяснимости и интерпретируемости результатов полученного решения множество V предполагается подмножеством модели протомультиверсальных чисел [4].

Заключение

Рассмотрена задача понижения размерности смыслового пространства, обоснована её актуальность. Предложены интерпретируемые (в рамках смыслового пространства) методы решения этой задачи на основе линейного автокодировщика с адаптивным коэффициентом обучения. Предложен метод расчёта адаптивного коэффициента обучения. На основе предложенного метода реализован алгоритм и проведено его тестирование в серии вычислительных экспериментов, результаты которых подтвердили корректность предложенного метода.

1. Manin, Yu. I., Marcolli, M. Semantic spaces. Math. Comput. Sci. – 2016. – 10, no. 4, – 459–477 pp.
2. Ivashenko, V. Approaches to the Study of Semantic Space and Integrated Logical Inference Models Using Similarity, Difference and Other Measures = Подходы к исследованию семантического пространства и моделей интегрированного логического вывода с использованием мер сходства, различия и других мер / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 97–112.
3. Ивашенко, В. П. Метрика, топология и персистентность в смысловом пространстве / В. П. Ивашенко // Информационные технологии и системы 2024 (ИТС 2024). – Минск: БГУИР, 2024. – С. 53–54.
4. Ivashenka, V. Semantic Space: between Continuity and Infinity / V. Ivashenka // Pattern Recognition and Information Processing (PRIP'2025) : Proceedings of the 17th International Conference, 16–18 September 2025, Minsk, Belarus / United institute of informatics problems of the National academy of sciences of Belarus. – Minsk : UIIP NASB, 2025. – P. 108–112.
5. Baldi, P. Deep Learning in Science. Cambridge: Cambridge University Press. – 2021. – 366 pp.
6. Nasir, S. A Survey on Multidimensional Scaling // S. Nasir, N. Haewoon, H. Mian, B. D. Muhammad, / ACM Computing Surveys. – 2018. – 1, no. 1, – 26 pp.
7. Головко, В. А. Нейросетевые технологии обработки данных : учеб. пособие / В. А.Головко, В. В. Краснопрошин. – Минск : БГУ, 2017. – 263 с.