

1. Предметная область

1. Ассоциативный эксперимент

1. Основные определения

2. Типы экспериментов

3. Аналоги

4. Ассоциативно-вербальный эксперимент

2. Машинное обучение

1. Типы решаемых задач

2. Основные алгоритмы классификации

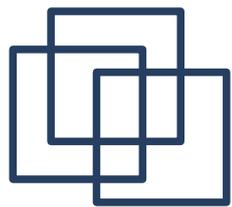
3. Основные алгоритмы кластеризации

4. Нейросети

2. Анализ данных

3. Построение математической модели

4. Список литературы

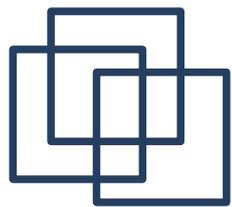


Нейросетевая модель вербального сознания. Предметная область. Ассоциативный эксперимент. Определения

Ассоциация (Павлов И.) - закономерная связь между отдельными событиями, фактами, предметами или явлениями, отражёнными в сознании и закреплёнными в памяти.

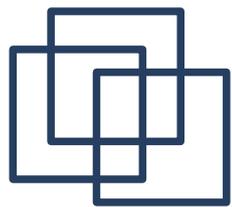
Появление одного (**стимула**) влечет по за собой появление другого (**реакции**)

YELLOW	BLUE	ORANGE
BLACK	RED	GREEN
PURPLE	YELLOW	RED
ORANGE	GREEN	BLACK
BLUE	RED	PURPLE
GREEN	BLUE	ORANGE



4 типа ассоциативных экспериментов:

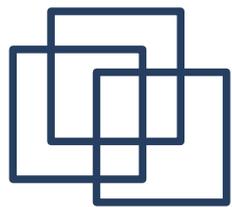
- **единичная свободная** (предлагается написать первое пришедшее на ум понятие);
- **единичная контролируемая** (в отличие от свободной, требуется написать определенный тип ассоциации (например, антоним или синоним));
- **множественные свободные** (респондент может использовать любое количество слов);
- **множественные контролируемые** (в отличие от предыдущей, накладываются ограничения на тип ответа)



Нейросетевая модель вербального сознания. Предметная область. Ассоциативный эксперимент. Типы экспериментов

Примеры:

- **единичная свободная:** *мать->дитя, турецкий->гамбит, зеленый->чай*
- **единичная контролируемая:** *синонимы: дитя->ребенок, веселый->радостный; антонимы: женщина->мужчина, высокий->низкий*
- **множественные свободные:** *мать->родная, женщина; турецкий->дворец, кофе*
- **множественные контролируемые:** *прилагательные: кофе->горький, черный; существительные: бутылка->минералки, Pepsi*



Нейросетевая модель вербального сознания.
Предметная область. Ассоциативный эксперимент.
Русский ассоциативный словарь

Тип: единственная свободная ассоциация

Период проведения: 1986-1997

Количество стимулов: 6624 (6577)

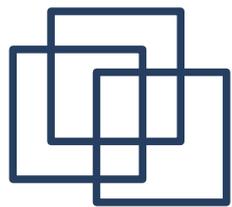
Количество анкет: 5000

Количество стимулов в каждой: 100

Количество респондентов: 11 000 (34 специальности)

Время на анкету: 7-10 минут

Объем словаря (различных пар стимул-реакция):
102516



Нейросетевая модель вербального сознания.

Предметная область. Машинное обучение.
Типы решаемых задач

- **Классификация (обучение с учителем)**

Техника, в ходе которой испытываемая система принудительно обучается с помощью примеров «стимул-реакция». Поиск функции, определяющей принадлежность объекта к определенному классу.

Методы: ИНН, SVM, random forest, boosting, байесовы сети, деревья решений, ансамбль классификаторов, скрытые марковские модели, регрессионный анализ, дискриминант Фишера

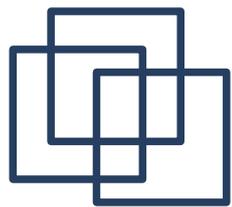
Приложения: компьютерное зрение, распознавание образов, речи, обработка естественного языка, кредитный скоринг, поисковые системы

- **Кластеризация (обучение без учителя)**

Техника для определения организации данных

Методы: k-средних, ближайших соседей, АГК, самоорганизующиеся карты, многомерное шкалирование

Приложения: Анализ социальных сетей, сегментация изображений, группирование объектов, снижение размерности



Постановка задачи

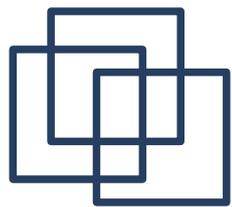
X - входные объекты (вектора)

T - ответы (истина)

Признаки: $f: X \rightarrow D_f$ $x = (f_1(x), f_2(x), \dots, f_m(x))$

Целевая ф-я: $y^*: X \rightarrow T$ $X^N = \{(x_1, t_1), (x_2, t_2), \dots, (x_N, t_N)\}$

Функция потерь: $E(w) = \frac{1}{2} \sum_{n=1}^N (y(x_n, w) - t_n)^2 \rightarrow \min_w$



Обучающая выборка: $X^L = \{x_1, x_2, \dots, x_l\}$

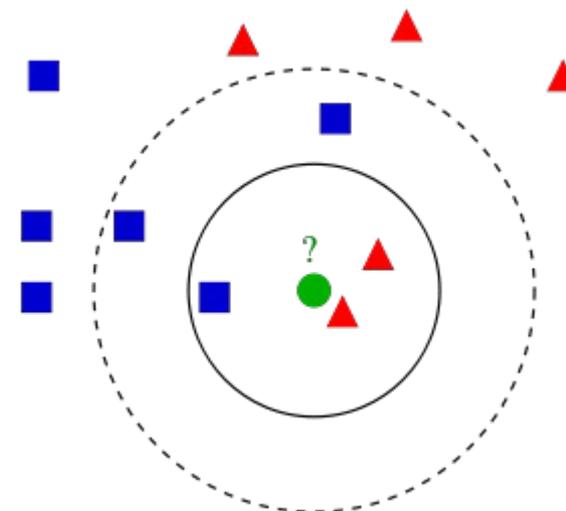
1. Матрица {объекты x признаки}:

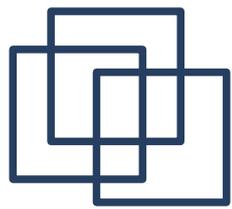
$$F = (f_j(x_i))_{l \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_l) & \dots & f_n(x_l) \end{pmatrix}$$

2. Матрица расстояний между объектами:

$$\Delta = \begin{pmatrix} \delta_{1,1} & \dots & \delta_{1,l} \\ \dots & \dots & \dots \\ \delta_{l,1} & \dots & \delta_{l,l} \end{pmatrix}$$

$$\|x_i - x_j\| = \delta_{i,j}$$

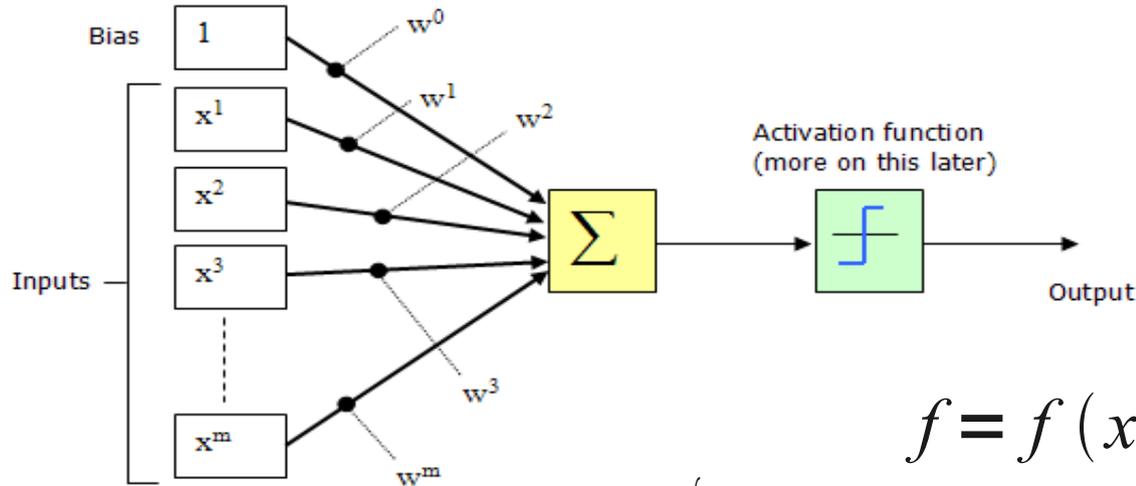




Нейросетевая модель вербального сознания.

Предметная область. Машинное обучение.
Нейронные сети. Модель нейрона

Маккалок, Питтс (1943):

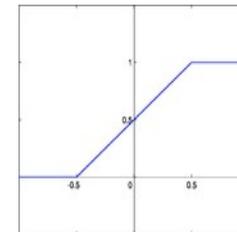


$$y = f(u)$$

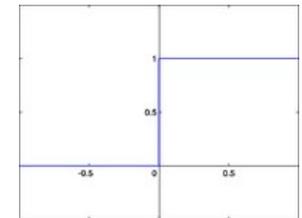
$$u = \sum_{i=1}^n w_i * x_i + w_0 * x_0$$

$f = f(x)$:

1. Линейная: $f(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x \geq 1 \\ x, & \text{else} \end{cases}$

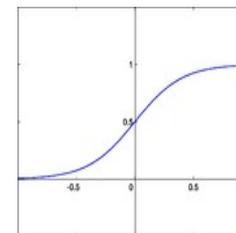


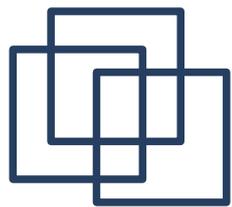
2. Хевисайда: $f(x) = \begin{cases} 1, & x \geq T \\ 0, & \text{else} \end{cases} \quad T = -w_0 * x_0$



3. Сигмоид: $f(x) = \frac{1}{(1 + \exp(-tx))}$

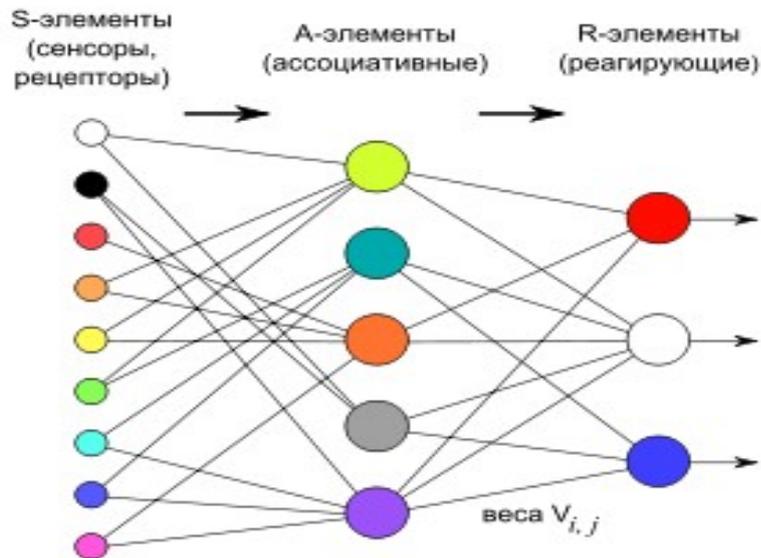
t-крутизна функции





Нейросетевая модель вербального сознания. Предметная область. Машинное обучение. Нейронные сети. Перцептрон

Розенблатт(1958):



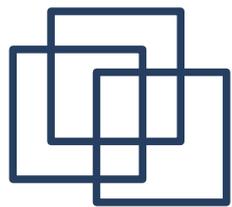
Алгоритмы обучения:

- Обратное распространение
- Метод коррекции ошибки

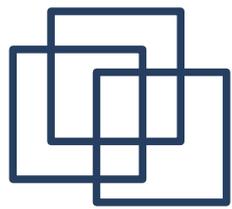
Сходимость (Новиков, 1962): $n_{max} = \left(\frac{D}{\delta}\right)^2$

$$D = \max \|x\|, x \in X^L$$

$$\langle x_i, \hat{w} \rangle y_i > \delta \quad i = 1..L, \delta > 0$$



- Части речи в реакциях: сущ. = ~53%, прилаг. = ~24% (стеммер mystem)
- «Сильные» стимулы: «лицо», «работа», «совет», «экономика», «труд», «наука», «страдание»
- «Сильные» реакции: «друг», «дурак», «работа», «вода», «ребенок», «дело», «есть»
- Имеется ядро основных понятий
- Пересечение по самым частым концептам с американским корпусом: «*food*», «*money*», «*water*», «*car*», «*good*»...

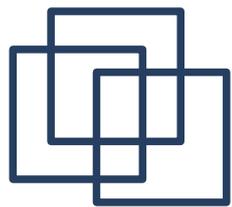


Графовая модель:

- Феномен «small-world» aka «6 степеней разделения» (для ненаправленного графа ассоциативной сети, WWW, энергосистемы США, сети научных соавторов)
- Функция степени вершины показательная:

$$P(k) \approx k^{-\gamma} \quad \gamma \in \{2..4\}$$

Аналогичный результат получен для американского корпуса



Нейросетевая модель вербального сознания. Русский ассоциативный эксперимент. Построение математической модели

1. Направленный->ненаправленный $S_{ij} = A_{ij} + A_{ji}$

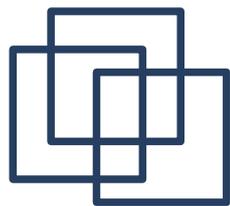
2. Учет косвенных связей:

$$\hat{S}_{ij} = -\log(S_{ik} S_{kl} \dots S_{nj}) = -\log S_{ik} - \log S_{kl} - \dots - \log S_{nj}$$

$S_{ik} S_{kl} \dots S_{nj}$ -вес кратчайшего пути из i в j

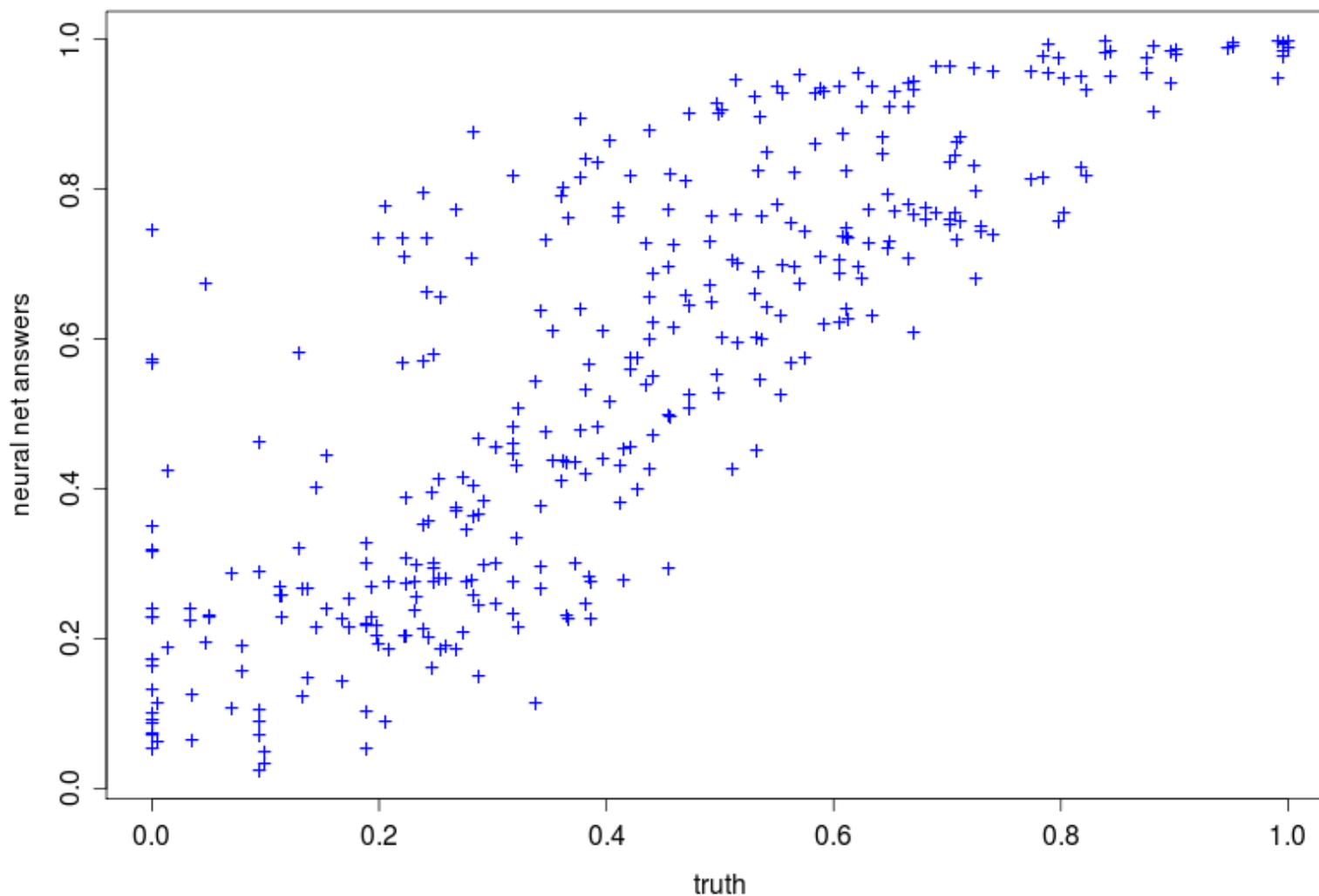
Матрица расстояний:

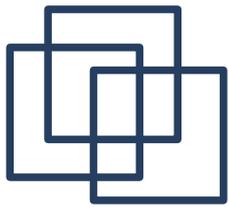
$$\Delta = \begin{pmatrix} \delta_{1,1} & \dots & \delta_{1,l} \\ \dots & \dots & \dots \\ \delta_{l,1} & \dots & \delta_{l,l} \end{pmatrix}$$



Нейросетевая модель вербального сознания. Русский ассоциативный эксперимент. Построение математической модели

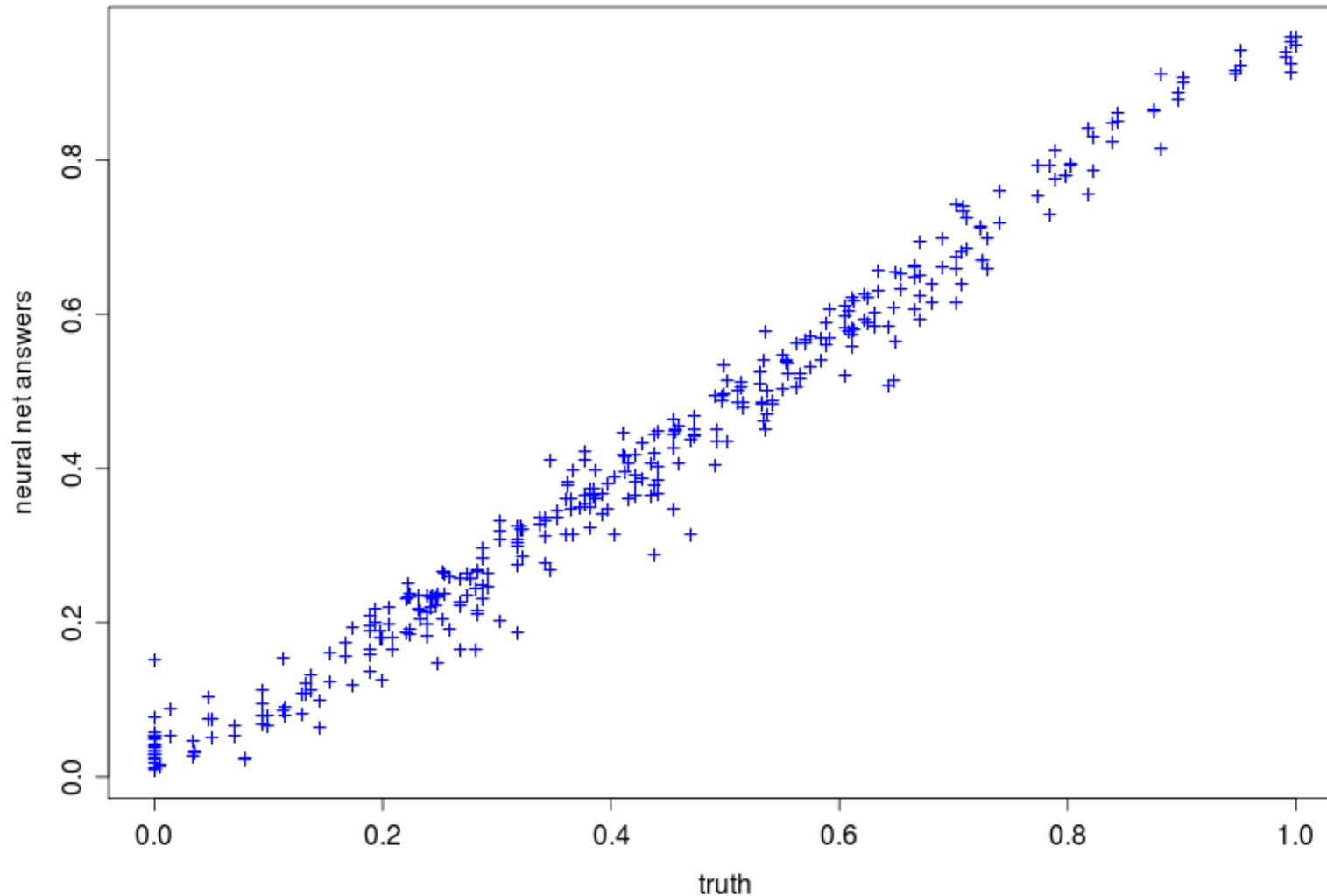
Бинаризация+перцептрон:

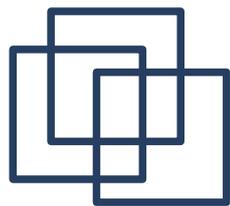




Нейросетевая модель вербального сознания. Русский ассоциативный эксперимент. Построение математической модели

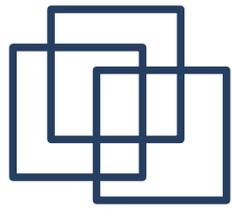
Многомерное шкалирование + перцептрон:





Нейросетевая модель вербального сознания. Русский ассоциативный эксперимент. Список литературы

- Deese, J. (1965). [The Structure of Associations in Language and Thought](#). Baltimore, MD: The Johns Hopkins Press.
- Jenkins, J. J., & Palermo, D. S. (1964). [Word Association Norms](#). Minn.:University of Minnesota Press
- De Groot, A. M. B. (1988). [Woordassociatienormen met reactietijden \(Word association norms with response times\)](#). Nederlands Tijdschrift voor de Psychologie (Dutch Journal of Psychology), 43, 280-296.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida [Word Association, Rhyme and Fragment Norms](#). <http://luna.cas.usf.edu/~nelson/>.
- De Deyne, S. & Storms, G. (2008). [Word associations: Network and semantic properties](#). Behavior Research Methods, 40, 213-231.
- Черкасова Г.(2008). [Русский сопоставительный ассоциативный словарь](#), Москва: Ияз РАН.
- Черкасова Г. (2004) [Формальная модель ассоциативного исследования](#).// Проблемы прикладной лингвистики. Выпуск 2. Сборник статей./ Отв. ред. Н.В.Васильева. Москва: «Азбуковник», 400с.
- Филиппович Ю., Дельфт Д., Черкасова Г. (2001). [Ассоциации информационных технологий: эксперимент на русском и французском языках.](#): Изд-во МГУМ.
- Черкасова Г. (2007) [Исследование динамики ассоциативно-вербальной модели языкового сознания русских](#).//Вопросы психолингвистики, 6.. Москва с.105-122
- Milgram, S. (1967, May). [The small-world problem](#). Psychology Today, 2, 60-67.
- Watts, D. J., & Strogatz, S. H. (1998, June 4). [Collective dynamics of 'small-world' networks](#). Nature, 393, 440-442.
- Albert, R., Jeong, H., & Barabási, A. L. (1999). [Diameter of the world wide web](#), Nature, 401,130-131
- Newman, M. E. J. (2001). [The structure of scientific collaboration networks](#). Proceedings of the National Academy of Sciences, 98, 404-409.
- Barabási, A. L., & Albert, R. (1999). [Emergence of scaling in random networks](#). Science, 286, 509-512.
- M. Steyvers, J. B. Tenenbaum (2005). [The large-scale structure of semantic networks: statistical analyses and a model of semantic growth](#):Cognitive Science, 29(1).
- А.В. Сиренко (2007). [Лингвокультурный тезаурус русского языка. Интеллектуальные технологии и системы. Сборник статей аспирантов и студентов](#). Выпуск 9 // Сост. и ред. Ю.Н. Филипповича. М.: Изд-во ООО «Эликс+».
- Steyvers, M., Shiffrin, R.M., & Nelson, D.L. (2004). [Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory](#). In A. Healy (Ed.), Experimental Cognitive Psychology and its Applications.
- Klauer, K. C., & Carroll, J. D. (1995). [Network models for scaling proximity data](#).In R. D. Luce, M. D'Zmura, D. D. Hoffman, G.



Спасибо за внимание!

Вопросы?