A Vector Field Approach to Lexical Semantics

Peter Wittek[1], Sándor Darányi[1], and Ying-Hsang Liu[2]

[1] University of Borås, Sweden
[2] Charles Sturt University, Australia

Abstract. We report work in progress on measuring "forces" underlying the semantic drift by comparing it with plate tectonics in geology. Based on a brief survey of energy as a key concept in machine learning, and the Aristotelian concept of potentiality vs. actuality allowing for the study of energy and dynamics in language, we propose a field approach to lexical analysis. Until evidence to the contrary, it was assumed that a classical field in physics is appropriate to model word semantics. The approach used the distributional hypothesis to statistically model word meaning. We do not address the modelling of sentence meaning here. The computability of a vector field for the indexing vocabulary of the Reuters-21578 test collection by an emergent self-organizing map suggests that energy minima as learnables in machine learning presuppose concepts as energy minima in cognition. Our finding needs to be confirmed by a systematic evaluation.

1 Introduction

In the context of Semantic Web dynamics [1], there is a growing body of literature about the semantic drift [2,3], the language-related version of abrupt parameter value changes in data mining called concept drifts [4,5,6,7]. By semantic drift we mean how the features of ontology concepts gradually change as their knowledge domain evolves. In what follows, we briefly outline a synoptic approach to the modeling of such a process from a QI perspective which departs from Aristotle and results in a vector field of word meaning. The evolution of this observable structure goes back to the underlying non-observable dynamics of concepts.

In the second section of this paper, we discuss how, in several disciplines, researchers have come up with more or less the same model over time, the combination of a continuous and a discrete plane to describe "fluid" content or experience being shaped into recognizable form for communication. The hidden layer of this construct is always field-like, sampled to build the discrete observable part. This biplanar construct can be compared to Aristotle's stance about potentiality versus actuality, including a distinction between dynamics and energy. Such a distinction, on the other hand, helps to spell out our working hypothesis.

In Section 3, we look at how machine learning uses energy minima on a potential surface to model both the goal state of learning, and concepts to be learnt for document categorization. Based on our past work, in Section 4 these ideas are combined to convert the vocabulary of a test collection of economic news into a vector field. After the experiment design, Section 5 describes the results and sets the scene for a systematic evaluation, not dealt with here. Section 6 brings us to our conclusions and refers to future work.

2 A tale of two planes

We find the tradition of using a combination of two planes to describe a phenomenon in several disciplines. Our first example is the general practice of evaluating the effectiveness of information retrieval and text categorization models by measures like recall, precision, accuracy, and many more [8]. For example there is ongoing work to build semantic spaces from distributional vs. compositional semantics [9,10], representing both word and sentence meaning as locations in high-dimensional space where for phrase or sentence component binding, recursive matrix-vector spaces [11], the tensor product [12,13,14], or circular holographic reduced representation are used [15]. In these models, the representation of semantic content in documents is compared to an ideal state of language use, provided by the human standards of interpretation inherent in the evaluation method [16]. Using geometry or probability as a vehicle of meaning, i.e. building a new medium of language, aims at maximizing similarity between the human standard and its statistical reconstruction. This hypothetic original, a correlate of spoken language called a mental state or internal state in neuroscience [17], recalls the "language of thought hypothesis" in philosophy [18], also called *mentalese*. What we can observe as a joint element in the above is that whereas language as a mental phenomenon is assumed to be continuous, its uttered or mathematically modelled representations are discrete.

The same duplicity returns as "hidden metaphysics" in traditional mentalist and more recent generalist-universalist theories about language: language is but a tool operated by something deeper – thought, reason, logic, cognition – which functions in line with biological-neurological mechanisms common to all human beings [19]. Moreover, a linguistic school of thought orthogonal to the above theories, called Neo-Humboldtian field theories of word meaning, goes back to the same dual model where discrete distributions of related content called lexical or semantic fields,

based on language use, are underpinned by the assumption of conceptual fields in the mind. Then, the lexical field of related words is only an outward manifestation of the underlying conceptual field so that the sum total of conceptual fields describes one's world view [20]. In yet another unrelated school of thought, Saussure's structural linguistics, language (*langue*) is a mental grammar with a rule set specifying ideal content pronunciation, whereas speech (*parole*) stands for the exemplification of those rules [21].

An important symptom of lexical fields is that regions of related content are separated by lexical gaps. These are nonexistent names for things where one could exist by rules of a particular language, and indicate possible conceptual distinctions not mapped to actual language use, such as mother's father (Swedish *morfar*) vs. father's father (Swedish *farfar*), both called *grandfather* in English, or father's brother (Swedish *farbror*) not distinguished from mother's brother (Swedish *morbror*), both called *uncle*. Such discontinuities of content play a prominent role in our working hypothesis.

2.1   Aristotle and QI

Apparently the assumption that products of the mind are continuous while their mapping to spoken language is discrete goes back ultimately to Aristotle's *Metaphysics*. In this, existence or reality is described as the sum total of two components, conceivable potentiality (*dynamis*) plus observable-measurable actuality (*energeia*). These are names for the latent vs. manifest capacity of existents to induce change.

As reviewed by Koznjak, (2007) [22], the first ones to link Aristotle and quantum mechanics were Bohm, (1951) [23] and Heisenberg, (1958) [24]. In the QI frame of thought, Aerts and Gabora, (2005) used the same insight [25]: a context-dependent property of a concept lends graded existence to it by weights spanning potential to actual existence because certain feature combinations are "less real", i.e. less typical for assessors. Therefore in our current thinking, existence consists of two layers, potentiality (a continuum) and actuality (a discrete distribution sampling the former). Importantly, one ascribes a field nature to mental experience because of the potentiality layer which we indirectly perceive by the actualized values of events.

The above have familiar repercussions in QI. For localized entities, the state of actuality correponds to one particular position of an observed particle (or a particular configuration of many particles), whereas potentiality means all possible locations or configurations in superposition.

Moreover reality is in the state of potentiality before and after observation, something that can be speculated about but not observed. On the other hand, actuality and the constant collapse of the wave function are the same. Phenomena pop in and out of existence: anything in the present is in the overlap between the last moment of the future and first moment of the past while being observed, but returns to uncertainty thereafter. This also means a link between potentiality as a continuous experience vs. actuality as its discrete mapping to real, objective existence. Last but not least, expanding on the implications, energy manifest in observed events (such as *parole*) must go back to dynamics latent in fields (here, *langue*).

2.2 Working hypothesis

In what follows, our working hypothesis will be that word meaning can be expressed as "energy" [26,27], because semantic content located in vector space generates a potential with energy minima on a potential surface. Such content constitutes regions with different semantic density [28] so that both concepts and categories as their combinations are modelled by the above minima. More importantly, using the concept of kinds [29], one can look into the dynamic "origins" of distributions of related content. To this end, one must add interpolation between located items of meaning such as by Gaussian blurring [30] – its role is to approximate a previous state of potentiality from observation as actuality. The result will show lexical gaps as metaphoric fault lines in plate tectonics, with new semantic content about to protrude.

3 Word semantics as a vector field: experiment design

Mathematical "energy" and machine learning (ML) are related, the latter often being based on minimizing a constrained multivariate function such as a loss function. Concepts in feature space "sit" at global energy minima, representing the cost of a classification decision as an energy minimizing process. This suggests that ML must identify concepts with such minima, and since potential energy in physics is carried by a field or a respective topological mapping, concepts naturally have something to do with energy as work capacity.

Our research problem below will be to model this "energetic" nature of language on a vector field, and relate the results to theoretical constraints set by distributional semantics. The novelty of this approach is that we integrate the energetic implications of ML algorithms with the

like nature of the raw material they process (Section 3.1). The solution we propose here is to use emergent self-organizing maps to generate an artificial semantic field. The space in this regard is a two dimensional surface, and the vector field associated with the points on the surface is a high-dimensional one. Each term in the corpus is associated with a neuron in the map, and additional neurons interpolate nearby terms; thus there are approximately five neurons per term (Section 4). We expect both lexical fields of meaning and lexical gaps between them to emerge in this model, making it useful for linguistic analysis (Section 5). Further we hope that combining this approach with earlier semantic models using the Hamiltonian of a quantum system, we will be able to come up with a dynamic model of language change (Section 6).

3.1 Energy-like objectives in machine learning

To underpin our working hypothesis, first we overview the metaphoric use of the energy concept in ML.

Supervised learning algorithms measure the difference between target labels and the predictions of the model being trained. The goal is to minimize the difference: in such a scenario, we may regard the objective as 'energy', and we look for a global minimum. This is not always the case, as error on the training sample does not necessarily imply a good generalization performance on unseen examples, as we know it from the theory of structural risk minimization. Hence, for instance, support vector machines do not fit this paradigm, but feedforward neural networks and certain types of boosting algorithms do. We spell out these two examples in more details.

In a feedforward neural network, multiple layers of neurons are connected by weighted edges. The last layer, the output layer, is the response of the network to stimulus. We study the squared error of the output node j for an input i:

$$E_{ij} = \frac{1}{2}(y_i - f_j(\mathbf{x}_i))^2, \tag{1}$$

where y_i is the correct label, and $f_j(\mathbf{x}_i)$ represents the output of node j. The energy of the system is the global error, the sum or average of these local error terms. We aim to minimize this objective function by adjusting the weights of the network. Since the response $f_j(\mathbf{x}_i)$ includes a sum of nonconvex functions, update schemes on gradient descent may not find the global minimum. Metaheuristics such as simulated annealing or quantum annealing can help moving on from local minima.

Boosting algorithms, if they do not include constraints on sparsity, that is, they do not consider the principles of structural risk minimization, translate to a gradient descent. These algorithms iteratively and sequentially train weak classifiers with respect to a distribution, and adding them to a final strong classifier. As they are added, they are weighted to reflect the accuracy of the weak learners. For example, AdaBoost is a minimization of a convex loss function over a convex set of functions [31]. The loss being minimized is the exponential loss

$$\sum_i e^{-y_i f(x_i)}, \tag{2}$$

and we are seeking a function

$$f(x) = \sum_t \alpha_t h_t(x), \tag{3}$$

where $h_t(x)$ is the output of a weak classifier t, and α_t is its weight in the strong classifier. Finding the optimum equals to a coordinate-wise gradient descent through a greedy iterative algorithm: it chooses the direction of steepest descent at a given step. Boosting problems even with nonconvex objective, translate to a quadratic unconstrained optimization problem of the following form:

$$\operatorname*{argmin}_{\mathbf{w}} \left(\frac{1}{L} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \left(\sum_{s=1}^{L} h_i(\mathbf{x}_s) h_j(\mathbf{x}_s) \right) - \frac{2}{L} \sum_{i=1}^{N} w_i \sum_{s=1}^{L} h_i(\mathbf{x}_s) y_s \right) \tag{4}$$

This formulation in turn is easily modeled by an Ising Hamiltonian.

Some unsupervised algorithms also seek a minimum on a high-dimensional surface, which, again, we may treat as a metaphor of energy. Examples include Hopfield networks, which map to an Ising Hamiltonian, or dynamic quantum clustering, where data instances are rolled along a potential surface to local minima.

In a Hopfield network, the array of neurons is fully connected, although neurons do not have self-loops [32]. This leads to $n(n-1)$ interconnections, with a w_{ij} weight on each. In this arrangement, the neurons transmit signals back and forth to each other in a closed feedback loop, eventually settling in stable states. The state s_i of a unit is either $+1$ or -1. It is activated by the following rule:

$$s_j = \begin{cases} +1 & \text{if } \sum_j w_{ij} s_j \geq \theta_i \\ -1 & \text{otherwise,} \end{cases} \tag{5}$$

where θ_i is a threshold value corresponding to the node. Using the weights and the states, Hopfield networks have a scalar value associated with each global state which we can call energy:

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i\ s_i \tag{6}$$

This value monotonically decreases as weights are updated to store patterns in the network, hence the global state will eventually settle in a minimum. The energy formulation is identical to an Ising model.

Dynamic quantum clustering is more direct in using energy as a metaphor [33]. It takes ψ as the ground solution for the generic Hamiltonian of the Schrödinger equation:

$$H\psi = (T + V(x))\psi = E_0\psi, \tag{7}$$

where H is the Hamiltonian, T is the kinetic energy, V is the potential energy, and E_0 is the ground energy level. The algorithm evolves the Hamiltonian to identify the clustering structure by tracking the expectation values of the position operator X: $\langle\psi(t)|x|\psi(t)\rangle = \int \psi(x,t)^* x \psi(x,t) dx$. The expectation values of the position operator obey their corresponding classical equations of motion, that is, the centre of each wave packet rolls towards the nearest minimum of the potential, according to Newton's law of motion.

These energy-type objective functions find a good fit with physical implementations of quantum optimization. For instance, adiabatic quantum computing finds the ground state of an Ising Hamiltonian, and small-scale demonstrations with boosting are promising [34]. The quantum analogue of Hopfield networks, quantum associative memories, define a Hamiltonian to retrieve elements from memory, albeit storage and optimization is unrelated to the energy function [35].

3.2 Implementing a potential field with lexical gaps

We are now in a position to return to our above working hypothesis with the following observations: in ML, the only reason why gradient descent and – in a more approximative fashion – simulated annealing algorithms work is that minima as learning goal states overlap with minima as concepts, that is, the nature of the learning algorithm and the phenomenon are identical. Put another way, the learning function is isomorphic with the semantic substratum it is supposed to identify, both belonging in

function space. Hence we regard concepts as attractors in a conceptual field, with lexemes as lexical attractors in its respective lexical field mapping [36,37]. This view is supported by neurosemantics where concrete noun representations are stored in a spectral fashion [38].

It is immediately clear that such a view can be generalized to features and feature agglomerations as attractors. On the other hand, it is no less evident that the dynamics of language cannot be conceived without the linguistic parallels of force, work and energy.

Let us revisit the Schrödinger equation in Equation 7. Taking the spectrum of the Hamilton operator H in a finite dimensional space, Wittek and Darányi (2011) conjectured that index terms are associated with a set of eigenvalues, giving them a spectral signature [26]. The eigenvalues corresponded to the different senses of the word, where a higher level energy state was more unlikely to be occupied.

Following a different train of thought, Darányi and Wittek (2012) studied the kinetic term of the Hamiltonian, T, to identify words with weights, and derive dynamics through Ehrenfest's theorem [27].

What has been missing so far is the potential term in the Hamiltonian, which is also the most complex one. We venture a step towards defining a potential field by interpolating the distributional semantic description of term vectors.

4 Emergent self-organizing maps for semantic interpolation

The vectors of the term space in a vector space model point to disjoint locations in a high-dimensional space. A field, on the other hand, is defined at all points in a space. To bridge this problem, we work on a two-dimensional surface. Some points on this surface correspond to terms; the corresponding term vector is the value of the vector field in that point. We interpolate the vector field in all the other points of the surface.

The mapping between points on the surface and the term vectors is done by training a self-organizing map. A self-organizing map is a two-dimensional grid of artificial neurons. Each neuron is associated with a weight vector that matches the dimension of the training data. We take an instance of the training data, find the closest weight vector, and pull it closer to the data instance. We also pull the weight vectors of nearby neurons closer to the data instance, with decreasing weight as we get further from the best matching unit. We repeat this procedure with every training instance. This consists one epoch. We repeat the same process

in the second epoch, but with a smaller neighbourhood radius, and a lower learning rate when adjusting the weight vectors. We continue training several epochs, until the network remains stationary. The resulting network reflects the local topology of the high-dimensional space [39].

An important condition for us is that the total number of neurons, N_n, is much larger than the number of terms, N_t. For instance, given 12,000 terms, we expect to train a map with 60,000 neurons or more. The superfluous neurons will be the interpolation in the points which are not directly associated with a term (Figure 1). Such maps are called emergent self-organizing maps [40], and they require highly parallel computational models [41].

By default, the weight vectors of the self-organizing maps are initialized by random vectors. We train the map until updates slow down. There is at least one matching neuron for each data point, but, since this is an emergent map, most neurons do not have a corresponding data point. What does the weight vector of such a neuron mean?
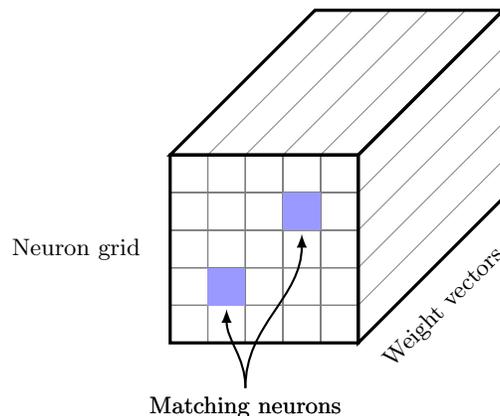

Fig. 1. A section of the two-dimensional surface of an emergent self-organizing map. At the end of the training, some neurons will correspond to index terms; these are shaded in the figure. Their corresponding weight vectors define the high-dimensional vector field at that point. Other neurons will interpolate the vector field between neurons that are associated with terms.

This weight vector is a form of interpolation. The weight vector was formed during the training as it was pulled by best matching neurons in its neighborhood. The weight vector is closest to its most nearby neuron that has a matching data point, but it is also somewhat similar to more remote but still close neurons with matching data points. This weight

vector fills the lexical gap between two data points, and corresponds to a gradually decreasing potential.

The grid is discrete, but the weight vectors in between neurons are easy to calculate by, for instance, a simple linear interpolation. Thus we can assign a high-dimensional weight vector to any point in the plane: we create a field.

Using a toroid map is advantageous as it avoids edge effects. The field is continuous in nature – a planar map would introduce an artificial discrete cut-off at the edges.

5 Results and evaluation

We used the vocabulary of the Reuters-21578 ML test collection of economic newswire items to generate an artificial lexical field. To measure lexical semantic relatedness, researchers typically compare the scores of words based on a taxonomic structure such as WordNet or Wikipedia with a gold standard as determined by correlations with human judgment [42,43]. For instance, an evaluation of topic coherence in topic modelling recruited users to score the topics with respect to their usefulness [44,45]. In this study, we resorted to a first evaluation of lexical gaps based on the structure of an emergent self-organizing map by "cherry picking" [46].

We used Lucene 3.6.2 to create an inverted index of the document collection. Terms were stemmed by the Porter stemmer, with those discarded that occurred less than three times or were in the top ten per cent of the most frequent ones. Thus we had 12,347 index terms, lying in an approximately twenty-thousand dimensional space.

We trained a toroid emergent self-organizing map of 336×205 dimensions using Somoclu [41]. The initial learning rate was 1.0, which decreased linearly over ten epochs to 0.1. The initial radius for the neighbourhood was a hundred neurons, and it also decreased linearly to one. The neighbourhood function was a noncompact Gaussian.

We studied the U-matrix of the map which depicts the Euclidean distance between the codebook vectors of neighbouring neurons, using the Databionics ESOM Tools for their visualization [40]. The global structure of the map is shown in Figure 2.

An important limitation of self-organizing maps is that they preserve the local topology but do not necessarily reflect the global topology of the high-dimensional space. If the initial weight vectors are not random but are primed with the global topology, then this problem is less severe.
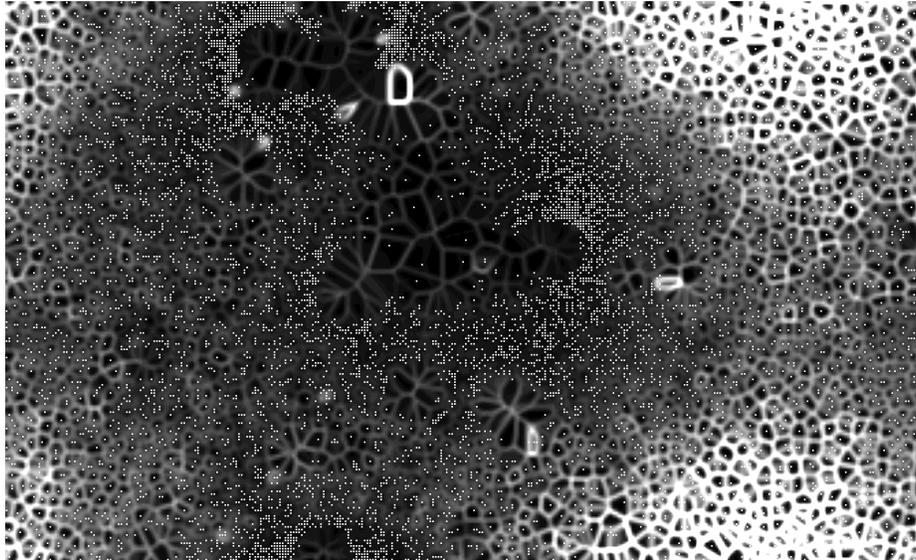
Fig. 2. U-matrix of the toroid emergent self-organizing map after ten epochs of training on the term space of Reuters-21578. The individual dots are neurons with a weight vector that match a term vector. All other neurons are interpolating the field.

We trained the map starting with random weights, hence local regions are topologically meaningful, but neighbouring regions in the map may be entirely unrelated.
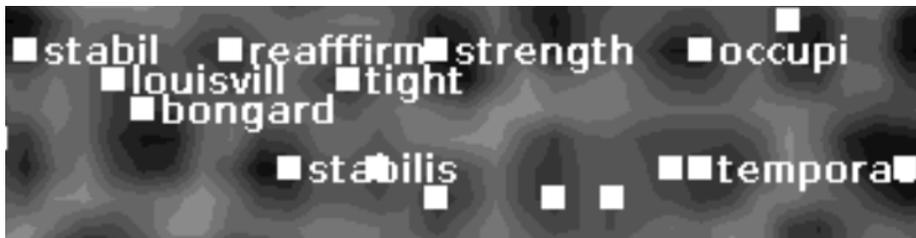
Fig. 3. A cropped section of the U-matrix with best matching units and labels, showing a tight cluster. Some labels are not displayed as they overlapped with others.

Figure 3 shows a tightly bundled group of terms. The gap between these words, based on the corpus, is small. The terms in this group, including ones that are not plotted in the figure, are: *bongard*, *consign*, *ita*, *louisvill*, *occupi*, *reafffirm* (with this misspelling), *stabil*, *stabilis*, *strength*,

temporao, tight. Some are clearly related, for others, we need to look at the corpus for justification. The expression Bongard appears in two senses: the corrupt head of a bank, and as the name of a brokerage firm. ITA always refers to the International Tin Association, which was debating a pact extension at the time. *Temporao* is a kind of cocoa, firms trading it were listed on stock exchanges. Louisville as a location appeared frequently in the economic news typical in this test collection. The gaps are small between these words, which does not necessarily rule out the insertion of new words in the gaps, but based on the limited vocabulary a newswire, the lexical field represented by these expression appears to be covered.

Fig. 4. A cropped section of the U-matrix with best matching units and labels, showing large gaps between words.

Large gaps are also interesting to look at. Take these examples (Figure 4): *energet*, *exploit*, *garrison*, *nato*, *petroleo*. Apart from *energet* and *garrison*, these words are frequent, with over twenty occurrences in the collection each. The reason for their isolation is not because their corresponding term vectors do not contain entries. These words are related, but their best matching units were pulled in other directions, creating a tension in the lexical field. Over time, words labelling new content could be expected to emerge in such "red hot" topic zones where metaphoric fault lines separate cells containing terms displayed as white dots. Such fault lines manifest lexical gaps, indicating content discontinuities in the observable field.

6   Conclusions and future work

Based on a brief survey of energy as a key concept in machine learning, and the Aristotelian concept of potentiality vs. actuality allowing for the

study of energy and dynamics in language, we proposed a field approach to lexical analysis. Until evidence to the contrary, it was assumed that a classical field in physics is appropriate to model word semantics. The approach used the distributional hypothesis to statistically model word meaning. We did not address the modelling of sentence meaning here. The computability of a vector field for the indexing vocabulary of the Reuters-21578 test collection by an emergent self-organizing map proved that energy minima as learnables in machine learning presuppose concepts as energy minima in cognition.

Next we plan a detailed evaluation of vector fields as vehicles of word meaning. This will focus on two things: a genuine lexical corpus instead of Reuters for the measurement of semantic and concept drifts, and the analysis of different field modification scenarios, including the context-dependent change of meaning of specific words; the dislocation of such terms and how their positions evolve over time ("drifts"); and the emergence of new words ("lexical gap studies").

In other words, we plan to look at the dynamics (i.e. the tensions) underlying semantic and concept drifts by specifically looking at:

- Distributional similarity studies [47,48], generalized to an algebraic form [49], also applied to image content [50];
- New studies of word meaning in context [16];
- Neural models of lexical semantics, e.g. Ursino et al 2010 who suggest that during processing [51], it is regions and not single locations that become activated in the brain.

We also intend to combine earlier efforts to use the Hamiltonian of a quantum system to model linguistic changes.

7 Acknowledgement

References

1. Antoniou, G., d'Aquin, M., Pan, J.Z.: Semantic web dynamics. Web Semantics: Science, Services and Agents on the World Wide Web **9**(3) (2011) 245–246

2. Lauriston, A.: Criteria for measuring term recognition. In: Proceedings of EACL-95, 7th Conference of the European Chapter of the Association for Computational Linguistics. (March 1995) 17–22

3. Gulla, J.A., Solskinnsbakk, G., Myrseth, P., Haderlein, V., Cerrato, O.: Concept signatures and semantic drift. In: Web Information Systems and Technologies. Springer (2011) 101–113

4. Delany, S.J., Cunningham, P., Tsymbal, A., Coyle, L.: A case-based technique for tracking concept drift in spam filtering. Knowledge-Based Systems **18**(4) (2005) 187–195

5. Wang, S., Schlobach, S., Klein, M.: Concept drift and how to identify it. Web Semantics: Science, Services and Agents on the World Wide Web **9**(3) (2011) 247–265

6. Ross, G.J., Adams, N.M., Tasoulis, D.K., Hand, D.J.: Exponentially weighted moving average charts for detecting concept drift. Pattern Recognition Letters **33**(2) (2012) 191–198

7. Gonçalves Jr, P.M., Barros, R.S.M.d.: Rcd: A recurring concept drift framework. Pattern Recognition Letters **34**(9) (2013) 1018–1025

8. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research **37**(1) (2010) 141–188

9. Padó, S., Lapata, M.: Dependency-based construction of semantic space models. Computational Linguistics **33**(2) (2007) 161–199

10. Erk, K., Padó, S.: A structured vector space model for word meaning in context. In: Proceedings of EMNLP-08, 13th Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, USA (October 2008) 897–906

11. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of EMNLP-CoNLL-12, Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. (July 2012) 1201–1211

12. Baroni, M., Lenci, A.: Distributional memory: A general framework for corpus-based semantics. Computational Linguistics **36**(4) (2010) 673–721

13. Blacoe, W., Kashefi, E., Lapata, M.: A quantum-theoretic approach to distributional semantics. In: Proceedings of NAACL-HLT-13, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA (June 2013) 847–857

14. Grefenstette, E., Dinu, G., Zhang, Y.Z., Sadrzadeh, M., Baroni, M.: Multi-step regression learning for compositional distributional semantics. arXiv:1301.6939 (2013)

15. Cohen, T., Widdows, D., Schvaneveldt, R.W., Rindflesch, T.C.: Discovery at a distance: Farther journeys in predication space. In: Proceedings of BIBMW-12, IEEE International Conference on Bioinformatics and Biomedicine Workshops. (2012) 218–225

16. Erk, K., McCarthy, D., Gaylord, N.: Measuring word meaning in context. Computational Linguistics **39**(3) (2013) 511–554

17. Elman, J.L.: An alternative view of the mental lexicon. Trends in Cognitive Sciences **8**(7) (2004) 301–306

18. Fodor, J.A.: The Language of Thought. Volume 5. Harvard University Press (1975)

19. House, J.: Linguistic relativity and translation. Amsterdam Studies in the Theory and History of Linguistic Science **4** (2000) 69–88

20. Trier, J.: Das sprachliche feld. Neue Jahrbucher fur Wissenschaft und Jugendbildung **10** (1934) 428–449

21. De Saussure, F.: Course in General Linguistics. (1916)
22. Kožnjak, B.: Möglichkeit, wirklichkeit und quantenmechanik. Prolegomena **6**(2) (2007) 223–252
23. Bohm, D.: Quantum theory. Dover Publications (1989)
24. Heisenberg, W.: Physics and Philosophy: The Revolution of Modern Science. Harper & Row (1958)
25. Aerts, D., Gabora, L.: A theory of concepts and their combinations I: The structure of the sets of contexts and properties. Kybernetes **34**(1/2) (2005) 151–175
26. Wittek, P., Darányi, S.: Spectral composition of semantic spaces. In: Proceedings of QI-11, 5th International Quantum Interaction Symposium. (June 2011)
27. Darányi, S., Wittek, P.: Connecting the dots: Mass, energy, word meaning, and particle-wave duality. In: Proceedings of QI-12, 6th International Quantum Interaction Symposium, Paris, France (June 2012)
28. Mihalcea, R., Moldovan, D.I.: Word sense disambiguation based on semantic density. In: Proceedings of COLING-ACL, 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. (August 1998)
29. Melucci, M.: Initial specifications for the design of information retrieval systems based on quantum detector using kinds. In: Proceedings of QI-13, 7th International Quantum Interaction Symposium. Springer, Leicester, United Kingdom (July 2013) 59–70
30. Darányi, S., Wittek, P.: Demonstrating conceptual dynamics in an evolving text collection. Journal of the American Society for Information Science and Technology **64**(12) (2013) 2564–2572
31. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. Annals of Statistics **28**(2) (2000) 337–407
32. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences **79**(8) (1982) 2554–2558
33. Weinstein, M., Horn, D.: Dynamic quantum clustering: A method for visual exploration of structures in data. Physical Review E **80**(6) (2009) 066117
34. Neven, H., Denchev, V.S., Drew-Brook, M., Zhang, J., Macready, W.G., Rose, G.: Binary classification using hardware implementation of quantum annealing. In: Demonstrations at NIPS-09, 24th Annual Conference on Neural Information Processing Systems. (December 2009) 1–17
35. Trugenberger, C.A.: Probabilistic quantum memories. Physical Review Letters **87** (July 2001) 067901
36. Amit, D.J.: Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press (1992)
37. Falissard, B.: A thought experiment reconciling neuroscience and psychoanalysis. Journal of Physiology – Paris **105**(4) (2011) 201–206
38. Just, M.A., Cherkassky, V.L., Aryal, S., Mitchell, T.M.: A neurosemantic theory of concrete noun representation based on the underlying brain codes. PLoS ONE **5**(1) (2010) e8622
39. Kohonen, T.: Self-Organizing Maps. Springer (2001)
40. Ultsch, A., Mörchen, F.: ESOM-Maps: Tools for clustering, visualization, and classification with emergent SOM. Technical report, Data Bionics Research Group, University of Marburg (2005)
41. Wittek, P.: Somoclu: An efficient distributed library for self-organizing maps. arXiv:1305.1422 (May 2013)

42. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics **32**(1) (2006) 13–47

43. Zhang, Z., Gentile, A.L., Ciravegna, F.: Recent advances in methods of lexical semantic relatedness–a survey. Natural Language Engineering **19**(4) (2013) 411–479

44. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: Proceedings of NAACL-HLT-10, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Stroudsburg, PA, USA, Association for Computational Linguistics (June 2010) 100–108

45. Wittek, P., Ravenek, W.: Supporting the exploration of a corpus of 17th-century scholarly correspondences by topic modeling. In: Proceedings of SDH-11, Supporting Digital Humanities: Answering the Unaskable, Copenhagen, Denmark (November 2011)

46. Kievit-Kylar, B., Jones, M.N.: Visualizing multiple word similarity measures. Behavior Research Methods **44**(3) (2012) 656–674

47. Weeds, J., Weir, D.: Co-occurrence retrieval: A flexible framework for lexical distributional similarity. Computational Linguistics **31**(4) (2005) 439–475

48. Rohde, D.L., Gonnerman, L.M., Plaut, D.C.: An improved model of semantic similarity based on lexical co-occurrence. Communications of the ACM **8** (2006) 627–633

49. Clarke, D.: A context-theoretic framework for compositionality in distributional semantics. Computational Linguistics **38**(1) (2012) 41–71

50. Bruni, E., Uijlings, J., Baroni, M., Sebe, N.: Distributional semantics with eyes: Using image analysis to improve computational representations of word meaning. In: Proceedings of MM-12, 20th ACM International Conference on Multimedia. (October 2012) 1219–1228

51. Ursino, M., Cuppini, C., Magosso, E.: A computational model of the lexical-semantic system based on a grounded cognition approach. Embodied and Grounded Cognition **1**(221) (2010)