ORIGINAL PAPER

Recursive neural networks prediction of glass transition temperature from monomer structure: an application to acrylic and methacrylic polymers

Celia Duce · Alessio Micheli · Roberto Solaro · Antonina Starita · Maria Rosaria Tiné

Abstract   We propose a new method based on a Recursive Neural Network (RecNN) for predicting polymer properties from their structured molecular representations. Rec-NN allows for a completely novel approach to QSPR analysis by direct adaptive processing of molecular graphs. This model joins the representational power of structured domains with Neural Network ability to capture underlying complex relationships in the data by a process of training from examples. To this aim, a structured representation was designed for the modelling of polymer structures. The adopted representation can account also for average macromolecule characteristics, such as degree of polymerization, stereoregularity, comonomer distribution. To begin with, this model was applied to the prediction of the glass transition temperature of (meth)acrylic polymers with different degree of main chain tacticity. The results so far obtained indicate that the proposed representation of polymer structure can convey information on both the repeating unit structure and average polymer features. The ability of the proposed Rec-NN method of treating this structured representation makes this method more general and flexible with respect to standard literature methods. Moreover, the same model can handle at the same time the Tg of polymer samples present in only one tacticity form together with that of polymer with different stereoregularity.

Keywords   Recursive neural networks · Cheminformatics · Glass transition temperature · Stereogularity · QSPR

C. Duce · R. Solaro (✉) · M. R. Tiné
Department of Chemistry and Industrial Chemistry, University of Pisa, via Risorgimento 35,
56126 Pisa, Italy
e-mail: rosola@dcci.unipi.it

A. Micheli · A. Starita
Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy

1 Introduction

Material performances must fulfil the stringent requirements of the diverse application of modern industry. Copolymerization, chemical modification, and blending allow for the production of a wide array of innovative polymeric products with tailored characteristics. On the other hand, experimental testing of new materials is a very expensive and time-consuming process. Consequently, the development of predictive methods to assess the most promising candidates for specific applications has gained urgency.

In time significant efforts have been spent on the development of Quantitative Structure Property Relationships (QSPR's) to predict physical, chemical, biological, and technological properties of molecules. Traditional QSPR approaches, employing standard regression methods (from linear regression to standard neural network) take as input fixed-size numerical vectors. As a consequence, all molecules must be reduced to vectors of the same dimension through the selection and extraction of their significant structural information.

The glass transition temperature (Tg) is often used as a benchmark of new methods for the prediction of polymer properties. This choice is suggested by the availability of a large number of experimental data. Additionally, it is well known that the glass-rubber transition is of considerable technological significance. In fact, the Tg determines the utilization limits of rubbers and thermoplastic materials. With respect to the prediction of this property, two main approaches have been used: group additive property (GAP) methods and systems that use molecular descriptors.

The additivity principle implies that a property, when expressed per mole of substance, can be calculated by summation of either atomic, or group, or bond contributions. Van Krevelen applied the GAP theory to nearly 600 polymers to derive the Tg group contributions [1]. About 80% of the calculated Tg's differed less than 20 K from experimental values. GAP approaches can be applied only to polymers containing previously investigated structural groups. Koehler and Hopfinger [2,3] tried to overcome this limitation by combining the group additivity approach with molecular modelling. A four-regression correlation with 12.8 and 16.6 K standard deviation (S) was obtained for 12 poly(alkyl acrylate)s and 20 poly(alkyl methacrylate)s, respectively [2].

On a parallel research line, several authors developed different methods based on molecular descriptors. Perhaps the most widely referenced model is the one proposed by Bicerano [4]. Bicerano built a QSPR model that combined a weighted sum of structural descriptors (topological parameters derived from atom connectivity indices) along with the solubility descriptor of each polymer. Application of a linear regression procedure to 320 polymer produced a model with $S = 24.65$ K and 0.9749 correlation coefficient (R). However, the prediction power of the above methods was not validated by an external test set. Waegell and co-workers built a model that is a combination of atomistic simulation and classical QSPR methods [5,6]. This model is called EVM (Energy, Volume, Mass) because of the three descriptors used in the QSPR equation. When applied to a training set of 16 aliphatic acrylic and methacrylic polymers, the standard deviation for a 12 polymers test set was 13 K. A linear correlation between Tg and the logarithm of the

effective side chain diameter/length ratio was proposed by Gao and Harmond [7]. The regressions obtained for nine poly(alkyl methacrylate)s and for five poly(alkyl acrylate)s gave in both cases a standard deviation of 7 K. Katritzky and co-workers calculated structural descriptors for the middle unit of a three-repeat units string [8]. A five-parameter model applied to 88 polymers produced $R^2 = 0.754$ and $S = 32.9$ K. Almost the same data set (84 polymers) was employed by Garcia-Domenech and de Julián-Ortiz [9]. The best regression was obtained with 10 variables (all graph-theoretical indices) and molar Tg's as dependent variable. All the descriptors used were calculated for the molecular monomer structure. The mean regression error for the training set and the cross validation were 12.7 and 15%, respectively.

Joyce and co-workers used feed-forward neural networks for the Tg prediction of 360 homopolymers by using an encoding of the SMILES symbols for a vector-based monomer representation [10]. Their model predicted the Tg of a polymer test set covering a wide structure range with $S \approx 35$ K. Sumpter and Noid [11,12] used a vectorial representation of the polymer repeating unit, based on descriptors obtained by combining different approaches [2,13–15]. Tg values were calculated with $S \approx 9$ K by applying a feed-forward neural network to 320 polymers from 23 different classes. Mattioni and Jurs used descriptors of different types of either the monomer or the polymer repeating unit [16]. The best model was obtained by using topological descriptors of the repeat unit structure. This model, applied to 251 polymers, generated $S = 21.1$, 25.2, and 21.9 K for training, cross-validation, and prediction sets, respectively.

In spite this progress, the prediction of polymer properties starting from their structure is a challenge still open. As it can be easily seen from the above survey, the main aim of any of the reported methods was the prediction of polymer properties starting from the molecular 2D graph. However, the only way to describe input data by using standard regression tools is to provide numerical variables expressing molecular descriptors through measurements or calculations. Because of computational limitations, the calculations of many types of standard descriptors are not carried out directly for high molecular weight molecules. Moreover, these models cannot account for average polymer characteristics, which often have a direct effect on the target property.

Direct structure treatment enables to by-pass the limitations associated with the use of molecular descriptors. The central point of our analysis stems from the fact that molecules are not simply fixed-size vectors of numbers but they are more naturally described via a varying size structured representation. Differing from other approaches, the method we propose is based on a Recursive Neural Network (RecNN) that can directly input variable-size labelled structures, such are Directed Positional Acyclic Graphs (DPAG) and rooted trees.

In the present paper, we propose a general and flexible representation of macromolecular structures in terms of labelled trees. This representation, which allows for analyzing different polymer classes simply from the 2D graph of the repeating units, can be extended to deal uniformly with both homopolymers and copolymers. Moreover, information accounting for the average characteristics of the macromolecules (degree of polymerization, main chain tacticity, monomer distribution, molecular weight, etc.) can be introduced in the input data. In its first implementation, this method was used

for the prediction of the glass transition temperature of a set of about 170 acyclic polymer samples of different stereoregularity. This set included alkyl, aminoalkyl, cyanoalkyl, fluoroalkyl, hydroxyalkyl, nitratoalkyl, oxaalkyl, sulfinylalkyl, thialkyl, cyanooxaalkyl, cyano-thiaalkyl, fluorooxaalkyl, and hydroxyoxaalkyl (meth)acrylate polymers, poly(meth)acrylamides, and a few α- and β-substituted poly(meth)acrylics.

2 Method

The adopted recursive neural network is a generalization of the well-known feed-forward neural networks for dealing with variable-size structured data e.g. labelled trees and labelled DPAGs (Directed Positional Acyclic Graphs) [17]. In such structures, for each vertex (or *node*) a total order is defined on the edges leaving from it and a position is assigned to each edge. We assume a bounded out-degree and that each DPAG possesses a super-source, i.e. a vertex *s* such that every vertex in the graph can be reached by a directed path starting from *s*. *Labels* are tuples of variables attached to vertexes. In particular, we consider a sub-class of these structures, the *k-ary trees* (*trees* in the following), which are rooted positional trees with finite out-degree *k*, i.e. *k* is the maximum number of children for each node.

The RecNN exploits a recursive encoding process, which mimics the morphology of each input hierarchical structure (such as rooted trees). For each vertex of the input structure, the model computes a numerical code by using information of both the vertex label and, recursively, the code of the sub-graphs descending from the current vertex. This process computes a code of the whole molecular structure. The code is then mapped to the output property value. The encoding and mapping free parameters of the neural network are adapted to the task through the learning algorithm on the basis of the training examples. By this process, RecNN models a direct and adaptive relationship between molecular structures and target properties. In particular, the recursive model can learn an encoding of the input structured representation according to the given QSPR task. Hence, RecNN can automatically discover by learning the specific structural descriptors (numerical code) for the particular task to be solved. As a result, no *a priori* definition/calculation and/or selection of input properties are needed.

More detailed information on the adopted RecNN method can be found in references [17–20].

3 Polymer structure representation

The first goal of this work was to find a flexible representation of polymer structure. The basic idea was to extend the tree representation developed for functional compounds [21] to the polymer repeating unit. Each repeating unit was decomposed by using almost the same atomic groups, labels, and priority rules used for polyfunctional compounds (Appendix 6). The tree root was positioned on an additional super-source vertex (the group "Start") used to close the highest priority side of the repeating unit. The other end of the structural unit was capped by a "Stop" group with the only purpose of closing the molecule. Neither the Start nor the Stop groups affect the chemical

Fig. 1 Structure, chemical tree, input data file, and numerical labels associated with poly(methyl methacrylate). The input data file contains the dimension of the tree (number of nodes), the value of the target property, and the connection table. *Column 1* of the connection table reports the order number identifying a specific node, which corresponds to the group indicated in *column 2*; *columns 3–5* indicate the presence of a "child" identified by its order number (-1 means no child); *column 6* reports the numeric code of the associated numerical label

features of the groups they are linked to. The specific task of the Start group is to allocate the average characteristics of the macromolecules (molecular weight, stereo-regularity, polydispersity index, etc.) in its label. In the present study, the Start label conveyed information of the polymers tacticity as molar fraction of r dyads. The Start label is a vector orthonormal to all other labels. The Stop label is a null vector. The Start and the Stop groups were given the highest and lowest priority, respectively. This representation can be extended to deal with copolymers by rooting two or more repeating units in the super-source vertex that contains information on their distribution.

The chemical tree of poly(methyl methacrylate) and tree conversion to RecNN input data file is presented in Fig. 1 as an example. The input data file contains the tree dimension (number of nodes), the target property value, and the structure connection table. The numerical labels of the groups and the Start label are reported in the same figure. This is the representation of the input data used by RecNN to recursively read the structure, i.e. to access each sub-graph in the encoding process.

4 Experiments

The need for making choices is inherent in preparing any data set from different sources for correlative and/or predictive purpose. The use of data from several different sources

is a major cause of the considerable standard deviation found between measured and fitted Tg values whenever a general correlation is developed. Indeed, in some cases the reported glass transition temperatures for the same polymer differ by as much as 80 K. The Polymer Handbook [22] provided most polymer Tg's in the selected data set. These data are well-supported by other sources where some data missing from the Polymer Handbook can be found [1,4,5,7]. Information concerning the Tg dependence on polymer tacticity were taken from several sources [4,22–33].

The whole polymer set was divided into disjoint training and test sets for learning and validation processes. Structures were selected in order to make the test set representative of the various molecular size, topology, and functional groups of the investigated repeating units. A "guess" set was also built to test the prediction ability of the RecNN under challenging conditions. More specifically, guess set polymers have molecular features scarcely represented in the data set and/or highly uncertain targets. The predicted properties of this special set must be considered individually and not statistically.

Five different experiments were carried out. In order to have significant appraisal of the results, sixteen trials were carried out for each RecNN experiment and the results were averaged over the different trials. The initial connection weights used in each trial were set at random. Learning was stopped when the maximum error for each compound of the training set was below the preset value.

In the first experiment, the training and test sets contained 127 and 27 polymer samples; the guess set was constituted by 3 compounds. The target values ranged from 162 to 501 K in the training set and from 198 to 413 K in the test set. The lowest known Tg was chosen for each methacrylic polymer with linear alkyl side chain. As a result, the input Tg decreases with increasing the number of carbon atoms in the side chain. The Tg of poly(n-hexadecyl methacrylate) does not follow this pattern. Therefore, this polymer was included in the guess set. Poly(1H,1H-pentadecafluorooctyl acrylate), whose Tg deviates from the trend of poly(1H,1H-perfluoroalkyl acrylate)s, was also placed in the guess set. Poly(N-tertbutylmethacrylamide), the only methacrylamide polymer, was included in the guess set as well. The maximum error was set at 20 K. The investigated polymers were taken from references that in most cases did not provide any quantitative information on main chain tacticity [1,4,5,7,22]. In order to compare samples from different sources, a molar fraction of r dyads of 0.9, 0.6, and 0.1 was assigned to polymers qualitatively indicated as syndiotactic, atactic, and isotactic. A 70% content of r dyads was attributed to atactic polymethacrylates, by taking into account their larger tendency to syndiotactic propagation.

The same training, test, and guess sets were used in the second experiment; however, the learning was stopped when the maximum error for each compound was below 110 K.

In the third experiment, acrylic and methacrylic polymers containing a linear side chain with more than eight and twelve carbon atoms, respectively were moved out from the data set. Some additional samples with known tacticity [4,23–33] were added to the data set. For this compilation (Table 1), preference was given to reports where the polymer synthesis was specified, the tacticity was determined by NMR, and the Tg was measured by DSC. In all other cases, atactic polyacrylates and polymethacrylates

were assigned a molar fraction of r dyads of 0.6 and 0.7; 1 and 0 were given to syndiotactic and isotactic samples, respectively. The need for uniformity within the data set is responsible of the small difference in the molar fraction of r dyads attributed to isotactic and syndiotactic polymers in experiments 1–2 and 3–5. Consequently, the r dyad content of some samples arbitrarily changed in passing from experiments 1 and 2 to 3–5. The new splitting of the data set resulted in training, test, and guess sets containing 137, 26, and 4 samples. The compounds in the guess set were poly (*N-tert*butylmethacrylamide), as in the previous experiments; syndiotactic pol (*iso*propyl methacrylate), for which only estimated Tg is available [32]; poly(acrylic acid), and poly(*N-sec*butylacrylamide). Differently from all other samples, the last two polymers can form hydrogen bonds. Moreover, there is only one polyacid in the training set, poly(methacrylic acid), and a few polyamides with uncertain Tg values. The target values ranged from 197 to 501 K in the training set and from 208 to 433 K in the test set. The learning was stopped when the maximum error for each compound was below 20 K. The fourth and fifth experiments used the same training, test, and guess sets as the third experiment, but the maximum error was set at 60 and 110 K, respectively.

For each experiment, the complete list of training, test, and guess sets is given in Appendix 6, where the target Tg, the molar fraction of r dyads, the mean calculated output, and the relative standard deviation, σ_i, over the sixteen trials are reported for each polymer sample.

5 Results

During the training of the network, different results can be achieved by starting from diverse initial conditions. In fact, the connection weights of the RecNN model are initialized at random because of the use of a gradient-based technique to solve a least mean square problem. This is a general phenomenon for standard neural networks [34] and other local-search optimization algorithms. Accordingly, sixteen trials were carried out for the RecNN simulations and the results were averaged over the different trials. Specifically, the mean and maximum absolute errors, the correlation coefficient (R) and the standard deviation (S) were obtained by an ensemble averaging method (Table 2). The number of RecNN hidden units (HU) of each experiment is also reported together with the number of samples (N) and the percentage (PTP) of samples that are present in more than one tactic form in the training and test sets of the different experiments. As indicated in refs. 17 and 19, the hidden units are the recursive neural units that compute the values of the encoding function for each input tree.

It is worth noting that a naive approach based on the selection of the best results over the various trials can lead to an unsatisfactory and unreliable estimation of the model performance. Moreover, this practice discards potentially useful information on the model behaviour, which is stored in the discarded regression estimates. The use of a basic ensemble method avoids these problems while offering an improved regression estimate.

Table 1 Names, Tg values, molar fraction of syndiotactic dyads (r), and relevant references of the investigated polymers

Name	Molar fraction of r dyads	Tg (K)	Reference
Poly(methyl acrylate)	0.6	281	[33]
	0	283	[32]
Poly(ethyl acrylate)	0.6	249	[33]
	0	248	[32]
Poly(isopropyl acrylate)	1	270	[4]
	0.6	267	[33]
	0	262	[32]
Poly(sec-butyl acrylate)	1	253	[22]
	0.6	251	[33]
	0	250	[32]
Poly(methyl α-chloroacrylate)	1	450	[4]
	0.75	424	[31]
	0.71	419	[31]
	0.59	411	[31]
	0.52	409	[31]
	0.35	380	[31]
	0	358	[4]
Poly(ethyl α-chloroacrylate)	1	404	[23]
	0.8	377	[31]
	0.71	367	[31]
	0.54	356	[31]
	0.27	325	[31]
	0.16	320	[31]
	0	308	[23]
Poly(i-propyl α-chloroacrylate)	1	409	[4]
	0.87	402	[31]
	0.64	383	[31]
	0.36	369	[31]
	0.34	366	[31]
	0.05	343	[31]
	0	341	[4]
Poly(methyl methacrylate)	0.99	403	[26,27]
	0.96	396	[25]
	0.83	388	[26]
	0.79	382	[26]
	0.74	378	[26]
	0.64	367	[26]
	0.01	328	[4,26]
Poly(ethyl methacrylate)	1	393	[23,32]
	0.85	359	[27]
	0.7	338	[32]
	0	281	[33]
Poly(isopropyl methacrylate)	1	412	[32]
	0.75	359	[27]
	0	307	[27]
Poly(isobutyl methacrylate)	1	393	[32]
	0.7	326	[33]
	0	281	[33]
Poly(*tert*butyl methacrylate)	0.75	391	[28]
	0.55	359.5	[28]
	0.10	350	[28]

Table 1 continued

Name	Molar fraction of r dyads	Tg (K)	Reference
Poly(2-hydroxyethyl methacrylate)	0.7	359	[22]
Poly(2-hydroxyethyl methacrylate)	0.2	311	[22]

The results obtained in experiment 1 are satisfactory with a mean error of about 2 and 14 K for the training and the test set, respectively. The standard deviation for the test set is 19 K, corresponding to 6.5% of the average experimental Tg (292 K). By taking into account the uncertainty of the target values, in experiment 2 the learning was stopped when the maximum error was below 110 K. It is worth noting that in this case the model is under-exploited since a very rough fitting is imposed. Indeed, this experiment was performed to empirically challenge our method and to gauge the noise and outliers of experimental data. As expected, the mean errors of both the training and the test sets are obviously higher than those of the previous experiment (Table 2). In this case, however, it was important to analyse the mean error for each sample in the training set. The polyacrylamides with branched side chains, poly(acrylic acid), and poly(methacrylic acid) showed absolute errors greater than the mean training error. Differently from others polymers, polyamides and polyacids can form hydrogen bonds; therefore, they are often excluded from the experimental data set [5,7]. Moreover, the Tg of strongly polar polymers is significantly affected by the sample moisture content.

When tacticity information was available, atactic and syndiotactic samples presented larger errors than isotactic ones. Nonetheless, in spite of the limited number of polymers present in the training set with more than one tacticity form, the RecNN correctly calculated the relative position of the Tg outputs for poly(*iso*propyl methacrylate) series in the test set of experiments 1 and 2.

In all cases, the absolute errors increased with the number of carbon atoms in the side chain; the largest absolute errors (all above the mean training error) correspond to polymers with long ester chains. Very likely, the RecNN could not find a single structure-Tg relationship for polymers with long side chains. Indeed, the RecNN had to deal with different trends within different polymer classes.

The effects of alkyl branching in polymers was noted by many researchers [1,4,7, 35–39]. The decrease of Tg with the length of side chains is widely believed to be a plasticising effect of the side chains [35]. The recoil takes place when the number of side-chain methylenes reaches 18 for poly(*n*-alkyl methacrylate)s, and 7 for poly(*n*-alkyl acrylate)s [7]. An increase of the brittle point of *n*-alkyl polymethacrylates and polyacrylates starting from poly(*n*-dodecyl methacrylate) and poly(*n*-octyl acrylate), respectively was reported by Reheberg and Fisher [36]. The authors pointed out, however, that the transitions of polymers starting from poly(*n*-dodecyl methacrylate) and poly(*n*-octyl acrylate) represent melting points of crystalline waxes.

Recent studies support the experimental observation that only the melting temperature can be measured for acrylate and methacrylate polymers containing long *n*-alkyl side chains [36]. The idea is that the side chains of different monomeric units

Table 2 Statistical parameters of experiments 1–5

Experiment	Maximum tolerance (K)	N	PTP(%)	HU	Training set				Test set			
					Mean absolute error	Maximum absolute error	R	S	Mean absolute error	Maximum absolute error	R	S
1	20	127 tr 27 ts	20 11	40	1.88	19.10	0.9985	3.56	13.57	55.86	0.948	18.90
2	110	127 tr 27 ts	20 11	11	14.83	89.58	0.949	20.87	20.84	78.66	0.897	27.48
3	20	137 tr 26 ts	34 42	35	2.41	15.46	0.9983	3.95	17.31	64.24	0.9299	22.45
4	60	137 tr 26 ts	34 42	15	8.26	42.17	0.9867	11.14	16.42	53.58	0.9385	21.07
5	110	137 tr 26 ts	34 42	9	14.68	91.43	0.9533	20.74	17.71	60.71	0.9206	23.84

R Correlation coefficient, S standard deviation, HU number of RecNN hidden units calculated as the average of the number of hidden units over 16 trials, N number of the samples, and PTP percent of polymers present with more than one tacticity in the training and test sets. All statistical parameters, except R, are expressed in K
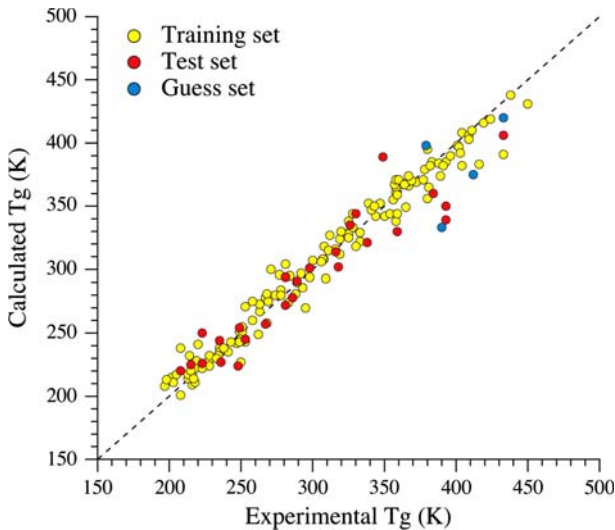
Fig. 2 Computed versus experimental Tg for polymers in experiment 4

from the same or from different macromolecules aggregate because the polar meth-acrylic main chain and the nonpolar alkyl side chains are immiscible. This leads to systems, which are more or less phase-separated, but only on a very short range. It has been shown [37–39] that for poly(n-alkyl methacrylate)s ($4 \leq CH_2 \leq 12$) and poly(n-alkyl acrylate)s ($4 \leq CH_2 \leq 10$), there are two coexisting relaxation processes with typical features of a dynamic glass transition: a polyethylene-like glass transition (α_{PE}) at lower temperature and the conventional glass transition (α) at higher temperature. The α_{PE} temperature increases with the methylene number, contrary to the α-process. Consequently, α_{PE} approaches α with increasing methylene number. In poly(dodecyl methacrylate) and poly(decyl acrylate) the two relaxation temperatures are similar and a single relaxation process is observed [37,38]. In the higher members, Tg(α_{PE}) occurs slightly below the crystallization onset, thus hindering the observation of main chain Tg that should occur at much lower temperature [39].

In agreement with the results of experiments 1 and 2 and by taking into account the above indications, acrylic and methacrylic polymers with more than 8 and 12 side chain methylenes were moved out of the data set in experiments 3–5. Long side chain poly(N-alkylacrylamide)s and polyacrylates with long perfluorinated side chains were taken out as well. On the other hand, the number of polymers present with different r dyad content almost doubled in going from the data sets of experiments 1 and 2 to those of experiments 3–5. As a result, RecNN efforts were made more difficult by the number of information that it was forced to learn. In other words, the RecNN had to handle a larger number of different Tg's associated to the same tree and set apart only by a parameter in the supersource label. Figure 2 presents a plot of computed *vs* experimental Tg for polymers in the data set of experiment 4.
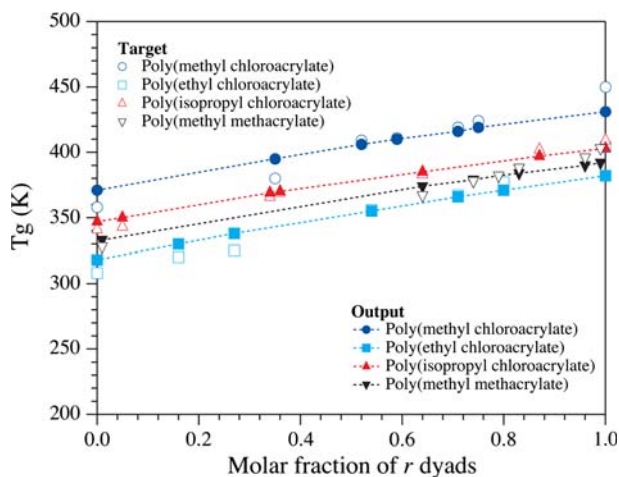
Fig. 3 Dependence of experimental and mean Tg output of training set polymers on *r* dyad content (Experiment 4)

The performances of test sets 4 and 5 were better than that of experiment 3 in which a high fitting level was assumed (Table 2). This result supports the hypothesis that a rather high noise persisted in the experimental data. The occurrence of noisy data discouraged any further fitting refinement. Indeed, the test set of experiment 4 gave the best results and experiments 4 and 5 presented a better mean training error/mean test error ratio than experiment 3. Punctual analysis of the training set outputs of both experiments 4 and 5, calculated at low fitting level, shows that some compounds i.e. poly(cyanoalkyl acrylate)s, poly(*N*-alkylamide)s, poly(methacrylic acid), poly(hydroxyalkyl methacrylate)s, and poly(fluoroalkyl methacrylate)s are outliers. This empirical observation is a RecNN indication of the target uncertainty for these compounds. The learning difficulty of the RecNN is reflected in the Tg prediction of the test set polymers of the same classes.

Analysis of the test set predictions of tactic polymer series is of paramount significance for verifying the real RecNN knowledge of the Tg-tacticity dependence. However, preliminary analysis of the training series of polymers with different molar fractions of *r* dyads is needed. The dependence of computed Tg in experiment 4 on the *r* dyad content is presented in Fig. 3 for the most interesting polymers. This experiment can be taken as a typical example of the RecNN performance in dealing with polymer tacticity. In all cases, the RecNN found an almost linear dependence of Tg on the molar fraction of *r* dyads. The slope of the regression fit decreased in going from experiment 3 to 5 (Fig. 4). However, the RecNN still learned the stereoregularity information even at the highest tolerance level.

The training outputs of polyacrylates have rather large errors and uncertainty, and the differences between the calculated Tg's of syndiotactic and isotactic forms are larger than the difference between experimental values. In fact, differently from polymethacrylates, the Tg of these polymers is almost independent of main chain stereoregularity. The learning of the poly(alkyl acrylate)s was obviously affected by the behaviour of the other polymers.
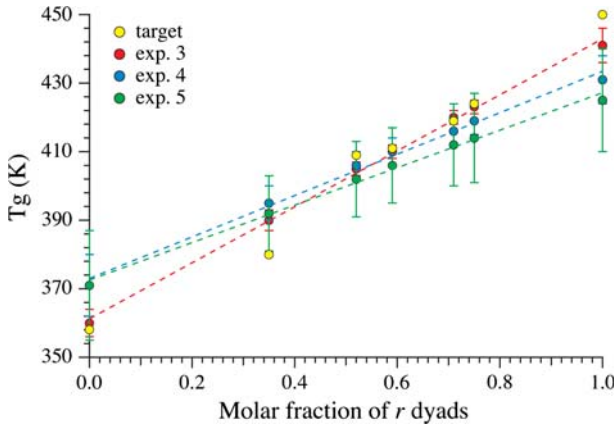
Fig. 4 Dependence of the mean Tg output from experiment 3–5 for poly(methyl chloroacrylate) samples with different molar fractions of *r* dyads
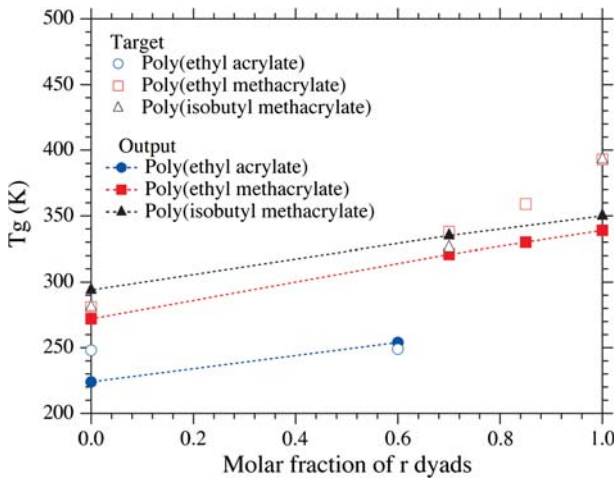
Fig. 5 Dependence of experimental and mean Tg output on *r* dyad content of test set polymers (Experiment 4)

Analysis of the test set results clearly shows that the RecNN correctly predicts the Tg-tacticity dependence of poly(ethyl acrylate), poly(ethyl methacrylate), and poly(isobutyl methacrylate) (Fig. 5). In particular, the output for these polymers correctly held the ranking of the target values, for each trial and in the averaged results. On the other hand, the RecNN performance was hampered in the poly(ethyl methacrylate) series; in this case, the Tg difference between syndiotactic and isotactic forms is 112 K, well above the difference found for any polymer series in the training set (a few Kelvin for poly(alkyl acrylate)s, 75 K for poly(methyl methacrylate), and from 68 to 96 K for poly(alkyl chloroacrylate)s).

Table 3 Target, mean output, and standard deviation, σ, obtained in the test set of experiment 4 for polymer samples with different r dyad content

Polymer sample	Molar fraction of r dyads	Target (K)	Reference[a]	Output (K)	σ (K)
Poly(ethyl acrylate)	0.60	249	[33]	254	20
	0	248	[32]	224	20
Poly(ethyl methacrylate)	1	393	[32]	339	20
	0.85	359	[27]	330	19
	0.70	324	[4]		
		338	[32]	321	19
		343	[6]		
		351	[5]		
		354	[5]		
	0	281	[33]	272	20
Poly(*iso*butyl methacrylate)	1	393	[32]	350	21
	0.7	321	[4]		
		340	[5]		
		326	[33]	335	23
	0	281	[33]	294	26

[a] When more target values were available for the same tacticity, only the values from references [32] and [33] were taken into account as the most reliable

It is worth noting that the experimental Tg for poly(ethyl methacrylate)s does not show a linear dependence on main chain tacticity. This behaviour is different from that of any other polymer in the training and test sets. Nonetheless, the mean errors in the predicted values of the test set are in the same range as literature data spread (30 K) for stereoregular polymers (Table 3). The steadiness of the mean output and the standard deviation for poly(ethyl methacrylate) throughout all learning levels (Experiments 3–5) can be taken as a RecNN hint of flawed experimental values. This indication is supported for instance by the identical reported Tg's of poly(ethyl methacrylate) and poly(*iso*butyl methacrylate), either isotactic or syndiotactic (Table 3). Moreover, the reported glass transition temperatures of syndiotactic poly(ethyl methacrylate) and poly(*iso*butyl methacrylate) are only estimated values [32]; the target values of isotactic poly(ethyl methacrylate) and poly(*iso*butyl methacrylate) were obtained by dilatometry; the samples stereoregularity was attributed by considering only the polymerisation method and it was not verified by NMR measurements [33].

Analysis of the guess set results highlighted some remarkable findings (Table 4). The mean calculated output for poly(acrylic acid) is excellent by taking into account the intricacy of the predictive work. Nevertheless, the high standard deviation indicates that the RecNN has a low confidence in the output, reflecting the lack of information on that compound. In the case of poly(*N-sec*butylacrylamide), the error and standard deviation reflect the difficulty met by the RecNN when learning polyacrylamides in the training set. By taking this into account, the poly(*N-tert*butylmethacrylamide) result is surprisingly good. Finally, the RecNN calculated a lower Tg value for isotactic poly(*iso*propyl methacrylate) than that reported in the literature [32].

The potential of the selected model to take into account the extent and type of stereoregularity of the polymer chains is of paramount importance because of the impact

Table 4 Targets, mean outputs, and standard deviations obtained for the guess set of experiments 3–5

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Output (K)	σ (K)	Output (K)	σ (K)	Output (K)	σ (K)
Poly(acrylic acid)	0.6	379	391	89	398	79	386	49
Poly(N-sec-butylacrylamide)	0.6	390	347	66	333	31	374	55
Poly(isopropyl methacrylate)	1.0	412	366	6	375	13	318	20
Poly(N-tert-butylmethacrylamide)	0.7	433	414	22	420	40	399	42

of this feature on many material properties. On the other hand, only a few authors try to correlate the Tg of polymers with their tacticity. Camelio's method [5,6] for Tg prediction makes calculation on a polymer segment of 20 repeat units to reduce the calculation time. However, a larger polymer segment (up to 50 units) should be considered to have a random distribution of repeating units [5]. Karasz and Mac Knight developed the most widely referenced model that rationalise the Tg dependence on the polymer stereoregularity [32]. They collected the available data for the glass transition temperature of vinyl polymers of the general formula $-[CH_2-CP(Q)]_m-$ and they observed that the steric configuration affects Tg only if P is different to Q and neither P nor Q are hydrogen [23,31,32].

One weakness of the reported methods is that for each polymer structure the used data sets must contain Tg values for different degrees of stereoregularity. On the other hand, our method allows for the contemporary treatment of polymers whose Tg is known at different tacticity degrees, together with atactic polymers and polymers whose Tg is known only at a given tacticity.

6 Conclusions

The proposed representation of polymer structure can convey information on both repeating unit structure and average polymer features. The capacity of the adopted RecNN to treat this structured representation makes our method more general and flexible with respect to standard literature procedures. In fact, when properly trained, the RecNN seems to be able to predict the dependence of the target property on either the polymer structure alone or both polymer structure and average properties. Moreover, the same model can handle polymers present in only one stereotactic form together with samples with a wide range of stereoregularity. The need for using different models to correlate different features to a defined target property is thus overcome. Analysis of the results clearly shows that the experimental Tg's of training set polymers are well reproduced. Moreover, the mean error of the predicted values in the test set is in the same range as literature data spread.

The RecNN uses a very small number of hidden units to solve the computational problem at the low fitting level imposed by the data noise. In other words, the RecNN considers the learning and prediction of the Tg data an easy computational task, if little significance is given to the most noisy data. With respect to other literature methods, the prediction ability of the proposed RecNN is rather good. Moreover, the reported analysis is useful to pick the most uncertain Tg data out of the data sets. This enable to exploit the RecNN, beside prediction, for data cleaning and data assessment purposes.

The RecNN potential was tested on an almost homogeneous class of homopolymers, but we plan to extend the data set to several classes of homopolymers, and then to copolymers. Indeed, the polymer representation was built bearing in mind the possibility of a potentially extension to deal with any polymer structure.

Moreover, analysis of the internal polymer representation computed by the RecNN could give a glimpse on the most relevant molecular features extracted by the RecNN for the Tg prediction. Indeed, this analysis proved very useful for the interpretation of

the RecNN learning process in the prediction of acyclic organic compound $\Delta_{solv}G°$ [21] and benzodiazepine activity [19].

Appendix 1: Representation of polymer structure

The repeating unit is partitioned in the following groups: CH_3, CH_2, C, H aliphatic, H aromatic, CF_3, CF_2, C=C, C≡C, OH, O, C=O, NH_2, NH, N, SH, S, S=O, CN, NO_3, F, Cl, Br, I. A numerical label is associated with each group. The labels discriminate among different groups of atoms and do not contain any physical–chemical information. The labels are represented by 27-bit vectors, with one or a few specific bits turned on and all others turned off. Sharing bits between different labels allows for representing similarity among chemical groups. Orthonormal vectors represent groups of different chemical nature. In particular:

- H aliphatic and H aromatic have orthonormal numerical labels.
- CH_3, CH_2, and C have similar numerical labels.
- CF_3 and CF_2 have similar numerical labels.
- N, NH_2, and NH have similar numerical labels.
- OH and O have orthonormal labels.
- SH and S have orthonormal labels.
- F, Cl, Br, and I have similar numerical labels.

The tree root is positioned on a super-source vertex (the group "Start") used to close the highest priority side of the repeating unit. The other end of the structural unit is capped by a "Stop" group with the only purpose of closing the molecule. Neither the Start nor the Stop groups affect the chemical features of the groups they are linked to. The Start label is a vector orthonormal to all other labels that conveys the average characteristics of macromolecules (molecular weight, stereoregularity, polidispersity index, etc.). The Stop label is a null vector. The Start and the Stop groups have the highest and lowest priority, respectively.

Priority scale

In aliphatic chains, the most substituted group has higher priority (C > CH_2 > CH_3 > H). A group containing a heteroatom has higher priority than any other group. In molecules with different heteroatoms, the C=O group has the highest priority. The priority decreases going to the right (N > O > F) and down (O > S; F > Cl > Br > I) in the periodic table. CF_2 and CF_3 have higher priority than CH_2 and CH_3, but lower than heteroatoms.

Edges order

The edges starting from a node are ordered according to priority rules. If two (or more) substituents in a node have the same priority, the groups along the substituent chains are ranked until a point of difference is reached.

Appendix 2: Results obtained in experiments 1–5

See Tables 5 and 6.

Table 5 (a) Training set, (b) test set and (c) guess set, experiments 1 and 2

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 1		Experiment 2	
			Out (K)	σ (K)	Out (K)	σ (K)
(a) Training set						
Poly(acrylic acid)	0.6	379	379	2	400	19
Poly(methyl acrylate) conv	0.6	283	283	4	301	31
Poly(ethyl acrylate) conv	0.6	249	248	3	260	14
Poly(ethyl acrylate) syn	0.9	249	260	5	277	15
Poly(ethyl acrylate) iso	0.1	248	237	6	239	12
Poly(propyl acrylate)	0.6	236	236	3	238	14
Poly(*iso*propyl acrylate) conv	0.6	267	266	3	279	11
Poly(*iso*propyl acrylate) syn	0.9	271	279	4	294	11
Poly(*iso*propyl acrylate) iso	0.1	262	255	7	260	12
Poly(*n*-butyl acrylate)	0.6	219	220	6	227	10
Poly(*iso*butyl acrylate)	0.6	249	249	2	262	21
Poly(*sec*butyl acrylate) conv	0.6	251	249	3	258	14
Poly(*sec*butyl acrylate) syn	0.9	253	260	4	274	14
Poly(*sec*butyl acrylate) iso	0.1	250	242	7	235	14
Poly(*n*-penthyl acrylate)	0.6	216	214	3	221	13
Poly(neopenthyl acrylate)	0.6	295	296	3	279	24
Poly(*n*-hexyl acrylate)	0.6	216	214	4	219	8
Poly(*n*-hepthyl acrylate)	0.6	213	212	5	220	13
Poly(2-hepthyl acrylate)	0.6	235	234	3	233	16
Poly(*n*-octyl acrylate)	0.6	208	209	5	218	10
Poly(*n*-nonyl acrylate)	0.6	215	216	2	229	12
Poly(*n*-dodecyl acrylate)	0.6	270	272	6	218	10
Poly(*n*-hexadecyl acrylate)	0.6	308	306	6	218	10
Poly(2-methylbutyl acrylate)	0.6	241	243	4	243	12
Poly(3-methylbutyl acrylate)	0.6	228	228	2	240	12
Poly(2-methyl-pentyl acrylate)	0.6	235	236	3	233	15
Poly(1,3-dimethylbutyl acrylate)	0.6	258	258	3	264	24
Poly(2-ethylhexyl acrylate)	0.6	223	223	4	226	14
Poly(2-methyl-7-ethyl-4-undecyl acrylate)	0.6	253	254	4	242	23
Poly(3-thiabutyl acrylate)	0.6	213	214	3	225	14
Poly(3-thiapentyl acrylate)	0.6	202	202	3	218	14
Poly(4-thiahexyl acrylate)	0.6	197	200	5	219	9
Poly(5-thiahexyl acrylate)	0.6	203	203	3	217	9
Poly(fluoromethyl acrylate)	0.6	288	287	3	293	33
Poly(2,2,2-trifluoroethyl acrylate)	0.6	263	263	3	260	24

Table 5 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 1		Experiment 2	
			Out (K)	σ (K)	Out (K)	σ (K)
Poly(1H,1H-pentafluoropropyl acrylate)	0.6	247	247	3	251	28
Poly(3-chloro-2,2-bis(chloromethyl)propyl acrylate)	0.6	319	319	2	319	34
Poly(1H,1H-heptafluorobutyl acrylate)	0.6	243	242	5	241	11
Poly(1H,1H,3H-hexafluorobutyl acrylate	0.6	251	251	2	243	8
Poly(1H,1H-nonafluoropentyl acrylate)	0.6	236	235	5	231	14
Poly(1H,1H,5H-octafluoropentyl acrylate)	0.6	238	238	3	242	14
Poly(cyanomethyl acrylate)	0.6	433	434	2	374	24
Poly(2-cyano*iso*propyl acrylate)	0.6	339	338	2	338	31
Poly(2-cyano*iso*butyl acrylate)	0.6	324	325	2	339	29
Poly(4-cyanobutyl acrylate)	0.6	233	234	4	242	18
Poly(2-cyanohexyl acrylate)	0.6	358	359	3	346	39
Poly(2-cyanoheptyl acrylate)	0.6	389	387	5	354	23
Poly(5-cyano-3-oxapentyl acrylate)	0.6	250	249	3	234	14
Poly(4-cyano-3-thiabutyl acrylate)	0.6	249	249	4	250	22
Poly(5-cyano-3-thiapentyl acrylate)	0.6	214	214	3	224	7
Poly(6-cyano-4-thiahexyl acrylate)	0.6	215	214	3	221	14
Poly(8-cyano-7-thiaoctyl acrylate)	0.6	223	223	2	229	15
Poly(2-methoxyethyl acrylate)	0.6	223	222	3	236	13
Poly(3-methoxypropyl acrylate)	0.6	198	199	2	225	14
Poly(3-ethoxypropyl acrylate)	0.6	218	218	3	220	7
Poly(3-methoxybutyl acrylate)	0.6	217	216	3	219	11
Poly(acrylamide)	0.6	438	438	2	430	23
Poly(N,*N*-dimethylacrylamide)	0.6	362	362	2	357	27
Poly(*N-iso*propylacrylamide)	0.6	358	359	3	364	21
Poly(*N-n*-butylacrylamide)	0.6	319	319	3	295	24
Poly(*N-sec*butylacrylamide)	0.6	390	390	3	354	20
Poly(*N-ter*butylacrylamide)	0.6	401	401	2	379	28
Poly(N,*N*-dibutylacrylamide)	0.6	333	333	2	320	32
Poly(*N*-(1-methylbutyl)acrylamide)	0.6	380	380	3	353	32
Poly(*N-n*-octylacrylamide)	0.6	220	220	3	241	21
Poly(*N-n*-octadecylacrylamide)	0.6	162	161	3	237	18
Poly(ethyl ethacrylate)	0.6	300	299	2	319	20
Poly[(methyl α-methoxycarbonylmethylacrylate	0.6	372	372	3	381	21
Poly(ethyl β-ethoxycarbonylmethacrylate)	0.6	325	324	3	333	26
Poly(*n*-hexyl β-hexyloxycarbonylmethacrylate)	0.6	269	270	2	264	22
Poly(methyl fluoroacrylate)	0.6	404	405	3	404	15
Poly(ethyl fluoromethacrylate)	0.6	316	316	2	325	17
Poly(methyl β-chloroacrylate)	0.6	416	416	3	363	34
Poly(ethyl α-chloroacrylate)	0.6	366	369	3	360	11
Poly(ethyl α-chloroacrylate) 10% iso	0.1	308	318	4	325	8
Poly(ethyl α-chloroacrylate) 100% syn	0.9	404	391	5	376	12
Poly(*n*-butyl α-chloroacrylate)	0.6	330	331	3	332	19
Poly(propyl α-chloroacrylate)	0.6	344	345	3	353	22
Poly(*n*-butyl α-cyanoacrylate)	0.6	358	356	3	340	22
Poly(*iso*propyl α-chloroacrylate)	0.6	363	364	3	368	18
Poly(methyl α-cyanoacrylate)	0.6	433	432	4	414	23
Poly(methacrylic acid)	0.7	501	501	2	457	33
Poly(methyl methacrylate)a	0.7	376	369	3	358	17
Poly(methyl methacrylate) iso	0.1	311	314	4	320	17

Table 5 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 1		Experiment 2	
			Out (K)	σ (K)	Out (K)	σ (K)
Poly(methyl methacrylate) syn	0.9	378	381	4	367	17
Poly(ethyl methacrylate) a	0.7	338	334	4	315	17
Poly(ethyl methacrylate) iso	0.1	285	285	4	275	15
Poly(ethyl methacrylate) syn	0.9	339	345	3	325	19
Poly(propyl methacrylate)	0.7	308	307	3	304	17
Poly(*n*-butyl methacrylate)	0.7	293	293	6	279	11
Poly(*n*-butyl methacrylate) iso	0.1	249	249	7	242	10
Poly(*iso*butyl methacrylate) a	0.7	321	323	2	328	16
Poly(*iso*butyl methacrylate) iso	0.1	281	278	4	290	21
Poly(*iso*butyl methacrylate) 80% syn	0.8	326	328	2	333	15
Poly(*ter*butyl methacrylate) a	0.7	377	364	3	358	18
Poly(*ter*butyl methacrylate) syn	0.9	387	380	3	368	20
Poly(2-hydroxyethyl methacrylate) 80% iso	0.2	311	314	5	308	16
Poly(2-ethylsulfinylethyl methacrylate)	0.7	298	297	5	284	25
Poly(2-nitratoethyl methacrylate)	0.7	328	329	2	338	24
Poly(2-ethylbutyl methacrylate)	0.7	284	284	3	289	26
(*b*) Test set						
Poly(*ter*butyl acrylate)	0.6	316	317	55	321	36
Poly(3-penthyl acrylate)	0.6	267	246	34	261	27
Poly(2-octyl acrylate)	0.6	228	216	2	229	12
Poly(tetradecyl acrylate)	0.6	297	301	14	218	10
Poly(2-ethylbutyl acrylate)	0.6	223	231	20	242	18
Poly(4-thiapentyl acrylate)	0.6	208	207	21	216	15
Poly(heptafluoro-2-propyl acrylate)	0.6	278	334	69	324	61
Poly(5,5,5-trifluoro-3-oxapentyl acrylate)	0.6	235	239	26	243	21
Poly(2-cyanoethyl acrylate)	0.6	277	290	31	302	36
Poly(2-cyanobutyl acrylate)	0.6	384	360	22	346	38
Poly(6-cyano-3-thiahexyl acrylate)	0.6	215	220	21	226	26
Poly(2-ethoxyethyl acrylate)	0.6	223	226	32	228	14
Poly(N,*N*-di*iso*propylacrylamide)	0.6	393	343	43	364	36
Poly(*N*-*n*-dodecylacrylamide)	0.6	198	211	46	239	20
Poly(butyl *β*-butoxycarbonylmethacrylate)	0.6	298	304	45	288	18
Poly(methyl fluoromethacrylate)	0.6	357	374	34	383	28
Poly(methyl chloroacrylate)	0.6	413	409	5	404	14
Poly(*sec*butyl chloroacrylate)	0.6	347	365	37	373	31
Poly(*iso*propyl methacrylate) a	0.7	354	339	32	345	26
Poly(*iso*propyl methacrylate) iso	0.1	300	289	37	308	26
Poly(*iso*propyl methacrylate) syn	0.9	358	351	32	354	28
Poly(*sec*butyl methacrylate)	0.7	333	329	37	323	14
Poly(*n*-octyl methacrylate)	0.7	253	224	35	238	18
Poly(3,3-dimethylbutyl methacrylate)	0.7	318	309	38	307	32
Poly(diethylaminoethyl methacrylate)	0.7	289	292	23	294	34
Poly(1H,1H,7H-dodecafluoroheptyl methacrylate)	0.7	286	271	33	299	24
Poly(2-hydroxypropyl methacrylate)	0.7	349	363	41	385	40
(*c*) Guess set						
Poly(*n*-hexadecyl methacrylate)	0.7	288	198	15	233	21
Poly(1H,1H-pentadecafluorooctyl acrylate)	0.6	256	239	35	235	12
Poly(N-*ter*butylmethacrylamide)	0.7	433	394	45	389	36

Table 6 (a) Training set, (b) test set and (c) guess set, experiments 3–5

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3 Out (K)	Experiment 3 σ (K)	Experiment 4 Out (K)	Experiment 4 σ (K)	Experiment 5 Out (K)	Experiment 5 σ (K)
(a) Training set								
Poly(methyl acrylate)	0.60	281	296	4	304	17	321	20
Poly(propyl acrylate)	0.00	283	270	5	274	13	289	25
Poly(isopropyl acrylate)	0.60	236	233	5	235	10	241	21
	1.00	271	282	6	300	13	325	24
Poly(butyl acrylate)	0.60	267	267	3	278	11	303	21
Poly(isobutyl acrylate)	0.00	262	250	5	249	11	273	19
Poly(secbutyl acrylate)	0.60	219	222	5	228	11	227	13
	0.60	249	250	4	251	16	257	23
	1.00	253	263	6	271	11	278	20
	0.60	251	251	3	251	7	258	17
	0.00	250	239	6	227	10	231	17
Poly(n-penthyl acrylate)	0.60	216	212	5	217	9	223	13
Poly(neopenthyl acrylate)	0.60	295	295	5	270	21	280	29
Poly(n-hexyl acrylate)	0.60	216	215	5	209	11	217	9
Poly(n-hepthyl acrylate)	0.60	213	214	5	214	10	221	13
Poly(2-hepthyl acrylate)	0.60	235	235	4	238	11	240	26
Poly(n-octyl acrylate)	0.60	208	205	6	201	9	216	9
Poly(2-octyl acrylate)	0.60	228	227	4	229	9	230	13
Poly(2-methylbutyl acrylate)	0.60	241	240	5	235	12	245	23
Poly(3-methylbutyl acrylate)	0.60	228	229	5	232	19	240	16
Poly(2-methyl-pentyl acrylate)	0.60	235	233	4	235	11	237	19
Poly(1,3-dimethylbutyl acrylate)	0.60	258	260	4	260	25	266	24
Poly(2-methyl-7-ethyl-4-undecyl acrylate)	0.60	253	254	4	243	20	239	23
Poly(3-thiabutyl acrylate)	0.60	213	214	5	217	10	230	16
Poly(3-thiapentyl acrylate)	0.60	202	204	6	215	10	223	12
Poly(4-thiahexyl acrylate)	0.60	197	201	5	208	10	218	8
Poly(5-thiahexyl acrylate)	0.60	203	205	4	211	12	217	9

Table 6 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Out (K)	σ (K)	Out (K)	σ (K)	Out (K)	σ (K)
Poly(fluoromethyl acrylate)	0.60	288	289	3	281	23	289	33
Poly(2,2,2-trifluoroethyl acrylate)	0.60	263	263	3	267	16	259	24
Poly(heptafluoro-2-propyl acrylate)	0.60	278	278	2	284	19	320	36
Poly(1H,1H-pentafluoropropyl acrylate)	0.60	247	247	5	242	16	233	12
Poly(3-chloro-2,2-bis(chloromethyl)propyl acrylate)	0.60	319	320	3	324	17	320	18
Poly(1H,1H-heptafluorobutyl acrylate)	0.60	243	243	4	243	20	238	20
Poly(1H,1H,3H-hexafluorobutyl acrylate)	0.60	251	250	3	255	17	248	17
Poly(1H,1H,5H-octafluoropentyl acrylate)	0.60	238	240	3	238	16	249	18
Poly(1H,1H-undecafluorohexyl acrylate)	0.60	234	234	3	232	12	237	17
Poly(2,2,3,3,5,5,5-heptafluoro-4-oxapentyl acrylate)	0.60	218	216	3	222	10	231	12
Poly(4,4,5,5-tetrafluoro-3-oxapentyl acrylate)	0.60	251	251	4	244	11	246	19
Poly(5,5,6,6,7,7,7-heptafluoro-3-oxaheptyl acrylate)	0.60	228	228	4	224	14	227	14
Poly(1H,1H-undecafluoro-4-oxaheptyl acrylate)	0.60	205	208	5	217	12	226	12
Poly(cyanomethyl acrylate)	0.60	433	433	4	391	17	342	18
Poly(2-cyanoethyl acrylate)	0.60	277	278	6	296	21	292	28
Poly(2-cyanoisopropyl acrylate)	0.60	339	339	3	352	18	316	28
Poly(2-cyanoisobutyl acrylate)	0.60	324	323	2	328	16	313	29
Poly(4-cyanobutyl acrylate)	0.60	233	235	7	230	18	242	18
Poly(2-cyanohexyl acrylate)	0.60	358	359	4	361	20	343	21
Poly(2-cyanoheptyl acrylate)	0.60	389	388	4	374	14	339	23
Poly(5-cyano-3-oxapentyl acrylate)	0.60	250	246	8	245	14	241	21
Poly(4-cyano-3-thiabutyl acrylate)	0.60	249	247	5	243	16	238	18
Poly(5-cyano-3-thiapentyl acrylate)	0.60	214	217	7	232	17	231	13
Poly(6-cyano-4-thiahexyl acrylate)	0.60	215	215	4	220	16	224	20
Poly(8-cyano-7-thiaoctyl acrylate)	0.60	223	224	4	226	10	225	18
Poly(2-methoxyethyl acrylate)	0.60	223	224	3	222	12	236	19
Poly(3-methoxypropyl acrylate)	0.60	198	200	4	213	14	226	16
Poly(3-ethoxypropyl acrylate)	0.60	218	214	6	211	7	219	9

Table 6 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Out (K)	σ (K)	Out (K)	σ (K)	Out (K)	σ (K)
Poly(3-methoxybutyl acrylate)	0.60	217	215	3	214	12	220	11
Poly(acrylamide)	0.60	438	437	4	438	14	418	34
Poly(N,N-dimethylacrylamide)	0.60	362	362	3	367	20	370	25
Poly(N-*iso*propylacrylamide)	0.60	358	360	6	365	14	361	19
Poly(N,N-di*iso*propylacrylamide)	0.60	393	393	3	385	17	362	29
Poly(N-butylacrylamide)	0.60	319	319	4	312	14	295	34
Poly(N-*ter*butylacrylamide)	0.60	401	400	4	398	13	379	30
Poly(N,N-dibutylacrylamide)	0.60	333	333	3	322	17	294	35
Poly(N-(1-methylbutyl)acrylamide)	0.60	380	376	5	356	13	323	34
Poly(N-*n*-octylacrylamide)	0.60	220	222	2	241	19	266	28
Poly(ethyl ethacrylate)	0.60	300	299	2	307	15	322	29
Poly[(methyl α-methoxycarbonyl-methylacrylate)	0.60	372	372	3	369	14	379	14
Poly(ethyl β-ethoxycarbonylmethacrylate)	0.70	325	326	3	328	14	335	20
Poly(*n*-hexyl β-hexyloxycarbonylmethacrylate)	0.70	269	271	4	275	11	283	15
Poly(methyl α-fluoroacrylate)	0.60	404	406	6	408	6	402	12
Poly(methyl α-fluoromethacrylate)	0.60	357	359	3	367	13	378	21
Poly(ethyl α-fluoromethacrylate)	0.60	316	316	3	316	15	325	21
Poly(methyl α-chloroacrylate)	0.00	358	360	4	371	9	371	16
	0.35	380	390	3	395	5	392	11
	0.52	409	405	2	406	4	402	11
	0.59	411	410	2	410	4	406	11
	0.71	419	420	2	416	4	412	12
	0.75	424	423	2	419	4	414	13
	1.00	450	441	5	431	7	425	15
Poly(methyl β-chloroacrylate)	0.60	416	413	3	383	15	366	27
Poly(ethyl α-chloroacrylate)	1.00	404	390	4	382	7	375	14
	0.80	377	376	2	371	6	367	10

Table 6 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Out (K)	σ (K)	Out (K)	σ (K)	Out (K)	σ (K)
	0.71	367	369	2	366	5	362	9
	0.54	356	356	2	355	4	353	8
	0.27	325	333	2	338	4	338	12
	0.16	320	323	2	330	4	331	14
	0.00	308	309	3	318	6	321	18
Poly(propyl α-chloroacrylate)	0.60	344	343	4	342	10	350	23
Poly(secbutyl α-chloroacrylate)	0.60	347	348	3	352	13	359	30
Poly(butyl α-cyanoacrylate)	0.60	358	358	3	338	15	320	22
Poly(isopropyl α-chloroacrylate)	0.00	341	342	4	347	8	349	15
	0.05	343	346	4	350	7	352	14
	0.34	366	366	2	369	5	369	11
	0.36	369	368	2	370	5	370	11
	0.64	383	386	2	385	5	386	10
	0.87	402	399	4	397	6	398	12
	1.00	409	407	5	403	6	404	14
Poly(methacrylic acid)	0.70	501	502	4	484	18	449	36
Poly(methyl methacrylate)	0.01	328	330	4	333	10	337	17
	0.64	367	373	2	374	7	372	11
	0.74	378	380	2	379	6	378	11
	0.79	382	383	1	382	6	380	12
	0.83	388	385	1	384	5	382	12
	0.96	396	394	2	390	6	388	13
	0.99	403	395	2	392	6	389	13
Poly(propyl methacrylate)	0.70	308	310	5	308	8	304	11
Poly(isopropyl methacrylate)	0.75	359	353	5	359	12	361	19
	0.00	307	312	6	308	12	318	25
Poly(butyl methacrylate)	0.70	293	290	6	286	11	277	12

Table 6 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Out (K)	σ (K)	Out (K)	σ (K)	Out (K)	σ (K)
Poly(butyl methacrylate) iso	0.00	249	248	7	243	9	248	22
Poly(secbutyl methacrylate)	0.70	333	334	5	329	12	315	16
Poly(terbutyl methacrylate)	0.75	391	383	3	382	9	370	24
	0.55	360	371	2	371	7	360	22
	0.10	350	347	3	342	11	335	23
Poly(n-pentyl methacrylate)	0.70	268	271	4	281	11	287	15
Poly(neopentyl methacrylate)	0.70	312	314	5	327	15	330	28
Poly(n-hexyl methacrylate)	0.70	268	266	6	258	12	270	16
Poly(3,5,5-trimethylhexyl methacrylate)	0.70	274	274	3	280	17	289	16
Poly(dimethylaminoethyl methacrylate)	0.70	292	292	4	297	16	305	16
Poly(2-terbutylaminoethyl methacrylate)	0.70	306	305	4	306	15	317	21
Poly(2-chloroethyl methacrylate)	0.70	365	365	2	349	22	336	25
Poly(2-bromoethyl methacrylate)	0.70	325	325	4	325	13	323	21
Poly(1,1,1-trifluoro-2-propyl methacrylate)	0.70	354	354	3	344	15	337	19
Poly(1H,1H-heptafluorobutyl methacrylate) syn	1.00	330	328	4	318	18	314	24
Poly(1H,1H,5H-octafluoropentyl methacrylate)	0.70	309	309	7	293	18	302	19
Poly(1H,1H,9H-hexadecafluorononyl methacrylate)	0.70	258	258	6	275	12	299	20
Poly(2-cyanoethyl methacrylate)	0.70	364	363	6	367	14	342	22
Poly(3-oxabutyl methacrylate)	0.70	289	289	3	290	19	292	19
Poly(3-oxa-5-hydroxypentyl methacrylate)	0.70	278	278	4	280	12	286	17
Poly(2-hydroxyethyl methacrylate)	0.70	359	352	5	344	13	335	24
Poly(2-hydroxyethyl methacrylate) 80% iso	0.20	311	318	5	315	12	313	25
Poly(2-ethylsulfinylethyl methacrylate)	0.70	298	295	6	294	12	291	13
Poly(2-nitratoethyl methacrylate)	0.70	328	329	3	344	13	331	26
Poly(2-ethylbutyl methacrylate)	0.70	284	284	3	295	16	292	20
(b) Test set								
Poly(terbutyl acrylate)	0.60	316	337	33	314	37	325	31
Poly(3-penthyl acrylate)	0.60	267	268	42	257	36	270	49

Table 6 continued

Polymer sample	Molar fraction of r dyads	Target Tg (K)	Experiment 3		Experiment 4		Experiment 5	
			Out (K)	σ (K)	Out (K)	σ (K)	Out (K)	σ (K)
Poly(2-ethylbutyl acrylate)	0.60	223	233	16	250	27	243	28
Poly(4-thiapentyl acrylate)	0.60	208	223	37	220	16	225	18
Poly(1H,1H-nonafluoropentyl acrylate)	0.60	236	232	21	227	15	227	12
Poly(5,5,5-trifluoro-3-oxapentyl acrylate)	0.60	235	234	35	244	37	226	18
Poly(2-cyanobutyl acrylate)	0.60	384	354	30	360	21	342	20
Poly(6-cyano-3-thiahexyl acrylate)	0.60	215	226	18	225	12	230	20
Poly(2-ethoxyethyl acrylate)	0.60	223	214	30	226	14	231	13
Poly(n-butyl β-butoxycarbonyl methacrylate)	0.70	298	302	14	301	20	289	15
Poly(n-butyl chloroacrylate)	0.60	330	355	22	344	11	333	26
Poly(methyl α-cyanoacrylate)	0.60	433	417	14	406	15	395	25
Poly(ethyl methacrylate)	1.00	393	341	23	339	20	332	23
	0.85	359	330	25	330	19	325	23
	0.70	338	318	27	321	19	318	22
	0.00	281	267	33	272	20	285	26
Poly(isobutyl methacrylate)	1.00	393	329	41	350	21	332	23
	0.70	354	310	39	335	23	318	19
	0.00	281	270	36	294	26	283	18
Poly(octyl methacrylate)	0.70	253	238	12	245	14	269	17
Poly(3,3-dimethylbutyl methacrylate)	0.70	318	310	42	302	22	299	19
Poly(ethyl acrylate)	0.60	249	267	36	254	20	267	29
	0.00	248	232	38	224	20	238	26
Poly(diethylaminoethyl methacrylate)	0.70	289	289	28	291	20	297	17
Poly(1H,1H,7H-dodecafluoroheptyl methacrylate)	0.70	286	271	16	278	13	300	20
Poly(2-hydroxypropyl methacrylate)	0.70	349	375	57	389	52	344	26
Poly(acrylic acid)	0.60	379	391	89	398	79	386	49
Poly(N-sec-butylacrylamide)	0.60	390	347	66	333	31	374	55
Poly(isopropyl methacrylate)	1.00	412	366	6	375	13	318	20
Poly(N-terbutylmethacryl amide)	0.70	433	414	22	420	40	399	42

References

1. D.W. Van Krevelen, *Properties of polymers-their estimation and correlation with chemical structure*, 2nd edn. (Elsevier, New York, 1976)
2. A.J. Hopfinger, M.G. Koehler, R.A. Pearlstein, J. Polym. Sci.: Part B: Polym. Phys. **26**, 2007–2028 (1988)
3. M.G. Koehler, A.J. Hopfinger, Polymer **30**, 116–126 (1989)
4. J. Bicerano, *Prediction of polymer properties. 3rd ed revised and expanded* (Marcel Dekker, New York, 2002)
5. P. Camelio, C.C. Cypcar, V. Lazzeri, B. Waegel, J. Polym. Sci.:Part A:Polym. Chem. **35**, 2579–2590 (1997)
6. C.C. Cypcar, P. Camelio, V. Lazzeri, L.J. Mathias, B. Waegel, Macromolecules **29**, 8954–8959 (1996)
7. H. Gao, J.P. Harmon, J. Appl. Polym. Sci. **64**(3), 507–517 (1997)
8. A.R. Katritzky, S. Sild, V.S. Lobanov, M. Karelson, J. Chem. Inf. Comput. Sci. **38**, 300–304 (1998)
9. R. Garcia-Domenech, J.V. Julián-Ortiz, J. Phys. Chem. B **106**, 1501–1507 (2002)
10. S.J. Joyce, D.J. Osguthorpe, J.A. Padgett, G.J. Price, J. Chem. Soc., Faraday Trans. **91**, 2491–2496 (1995)
11. B.G. Sumpter, D.W. Noid, J. Thermal. Anal. **46**, 833–851 (1996)
12. C.W. Ulmer II, D.A. Smith, B.G. Sumpter, D.I. Noid, Comput. Theor. Polym. Sci. **8**, 311 (1998)
13. A.A. Askadskii, Polym. Sci. USSR **9**, 471–487 (1966)
14. A.A. Askadskii, G.L. Slonimskii, Polym. Sci. USSR **13**, 2158–2160 (1971)
15. D. Porter, *Group interaction modeling of polymer properties* (Dekker, New York, 1995)
16. B.E. Mattioni, P.C. Jurs, J. Chem. Inf. Comput. Sci. **42**, 232–240 (2002)
17. A. Sperduti, A. Starita, IEEE. Trans. Neural. Networks **8**, 714–735 (1997)
18. A.M. Bianucci, A. Micheli, A. Sperduti, A. Starita, Appl. Int. J. **12**, 117–146 (2000)
19. A. Micheli, A. Sperduti, A. Starita, A.M. Bianucci, J. Chem. Inf. Comput. Sci. **41**, 202–218 (2001)
20. C. Duce, A. Micheli, R. Solaro, A. Starita, M.R. Tinè, in *Recursive neural networks for quantitative structure-property relationship analysis of polymers*, ed. by Simos Theodore, Maroulis George Lecture series on computer and computational sciences, vol. 4. (Brill Academic Publishers, Leiden, 2005), pp. 1546–1549
21. L. Bernazzani, C. Duce, A. Micheli, V. Mollica, A. Sperduti, A. Starita, M.R. Tiné, J. Chem. Inf. Model. **46**(5), 2030–2040 (2006)
22. J. Brandrup, E.H. Immergut (eds.), *Polymer handbook*, 3rd edn. (Wiley, New York, 1990)
23. B. Wesslen, R.W. Lenz, W.J. MacNight, F.E. Karasz, Macromolecules **4**(1), 24–26 (1971)
24. S. Bywater, P.M. Toporowsky, Polymer **13**, 94–96 (1972)
25. K. Ute, N. Miyatake, K. Hatada, Polymer **36**, 1415–1419 (1995)
26. A. Gourari, M. Bendaoud, C. Lacabanne, R.F. Boyer, J. Polym. Sci.: Polym. Phys. Ed. **23**, 889–916 (1985)
27. A.M. Walstrom, R. Subramanian, J.E. Long, J.E. McGrath, T.C. Ward, Polym. Prepr. **27**(2), 135–137 (1986)
28. E. Gipstein, E. Kiran, J.K. Gillham, Polym. Prepr. **13**(2), 1212–1217 (1972)
29. A.R. Katritzky, P. Rachwal, K.W. Law, M. Karelson, V.S. Lobanov, J. Chem. Inf. Comput. Sci. **36**, 879–884 (1996)
30. T. Kitayama, K. Ute, M. Yamamoto, N. Fujimoto, K. Hatada, Polym. J. **22**(5), 386–396 (1990)
31. G.R. Dever, F.E. Karasz, W.J. MacNight, J. Polym. Sci.: Polym. Chem. Ed. **13**, 2151–2179 (1975)
32. F.E. Karasz, W.J. MacKnight, Macromolecules **1**(6), 537–540 (1968)
33. J.A. Shetter, J. Polym. Sci. Part B **1**, 209–213 (1963)
34. S. Haykin, *Neural networks: A comprehensive foundation* (Prentice Hall, New Jersey, 1999)
35. M.P. Stevens, *Polymer chemistry*, 2nd edn. (Oxford University, New York, 1990), p. 80
36. C.E. Reheberg, C.H. Fisher, Ind. Eng. Chem. **40**(8), 1429–1433 (1948)
37. M. Beiner, E. Schröter, K. Hempel, S. Reissig, E. Donth, Macromolecules **32**, 6278–6282 (1999)
38. M. Beiner, H. Huth, Nat. Mater **2**, 595–599 (2003)
39. E. Hempel, H. Huth, M. Beiner, Thermochim. Acta. **403**, 105–114 (2003)