

Μελέτη σχέσης γλωσσολογικών χαρακτηριστικών, επιπέδων δημιουργικότητας και ηλικίας συγγραφέων για παραμύθια γραμμένα από παιδιά

στο πλαίσιο του μαθήματος

Προχωρημένα Θέματα Επεξεργασίας Φυσικής Γλώσσας

Αλέξιος Μανδαλιός

Υ.Δ. ΣΗΜΜΥ

A.M. 03002761

1 Εισαγωγή

Η φυσική γλώσσα περιέχει ρητή και άρρητη γνώση. Η ρητή γνώση είναι εύκολο να εξαχθεί από αυτόματα συστήματα επεξεργασίας στη μορφή χαρακτηριστικών κειμένου. Η παρούσα εργασία έχει ως στόχο να μελετήσει την εκπαίδευση μοντέλων που επιχειρούν τη μετάβαση από τη ρητή και εύκολα επεξεργάσιμη πληροφορία σε χρήσιμα συμπεράσματα για το κείμενο. Συγκεκριμένα, μελετάται η σχέση μεταξύ γλωσσολογικών χαρακτηριστικών κειμένων, ηλικίας των συγγραφέων, και αξιολογήσεων δημιουργικότητας. Στην ενότητα 2 αναπτύσσεται το θεωρητικό υπόβαθρο για την αξιολόγηση δημιουργικότητας στη μορφή μιας σύντομης βιβλιογραφικής ανασκόπησης. Στην ενότητα 3 παρουσιάζεται το dataset που θα χρησιμοποιηθεί. Στην ενότητα 4 περιγράφονται τα χαρακτηριστικά που εξάγονται από τα κείμενα και αντιπροσωπεύουν τη ρητή γνώση, και επιχειρείται η κατασκευή μοντέλων μηχανικής μάθησης που θα προβλέπουν χρήσιμα στοιχεία όπως βαθμός δημιουργικότητας και ηλικία συγγραφέα για κείμενα. Τέλος, στην ενότητα 5 καταγράφονται τα βασικά συμπεράσματα που εξάγονται από την εργασία.

2 Θεωρητικό υπόβαθρο

Ο στόχος του πεδίου επεξεργασίας φυσικής γλώσσας (Natural Language Processing/NLP) είναι η κατανόηση της ανθρώπινης γλώσσας. Υπάρχει πλήθος συστημάτων που εξάγουν χρήσιμα χαρακτηριστικά από ένα κείμενο, από τη δομή τους μέχρι τα νοήματα των λέξεων που περιέχουν. Ωστόσο, δεν είναι ξεκάθαρο το κατά πόσο αυτά τα συστήματα μπορούν να συνδέσουν τη ρητή πληροφορία που εξάγεται από το κείμενο με την άρρητη πληροφορία που αντιλαμβάνεται ο άνθρωπος.

Στην παρούσα εργασία μας ενδιαφέρει πόσο δημιουργικό είναι ένα κείμενο γραμμένο σε φυσική γλώσσα. Ο ορισμός της δημιουργικότητας διαφέρει για τις διάφορες σχολές σκέψης,

και η πεποίθηση ότι η δημιουργικότητα μπορεί να αξιολογηθεί με οποιονδήποτε αντικειμενικό τρόπο δεν είναι οικουμενικά αποδεκτή [Moz13]. Ακόμα και στις περιπτώσεις όπου θεωρείται ότι η δημιουργικότητα μπορεί να αξιολογηθεί, έχουν προταθεί διάφορες προσεγγίσεις, που περιλαμβάνουν:

- Θεωρίες δημιουργικότητας, που προσπαθούν να περιγράψουν τη διαδικασία με την οποία γεννιούνται οι δημιουργικές ιδέες. Όπως συμβαίνει συνήθως, είναι πιο δύσκολο να κατασκευαστεί ένα generative model παρά ένα predictive model.
- Χαρακτηριστικά που σχετίζονται με τη δημιουργικότητα, τα οποία συνδυάζονται σε δομές που καλούνται rubrics, με στόχο να οριοθετήσουν τη δημιουργικότητα σε κείμενα.
- Αξιολογήσεις ανθρώπων, συχνά απασχολούμενων στο πεδίο της γλώσσας με κάποιον τρόπο, γεγονός που τους καθιστά κατάλληλους να κρίνουν τη δημιουργική δουλειά των άλλων.

Οι τελευταίες δύο προσεγγίσεις μπορούν να συνδυαστούν, ώστε οι ειδικοί να παρέχουν την άποψή τους σε καθέναν από τους άξονες ενός συγκεκριμένου rubric. Ακόμα και αν δεχθούμε ότι η δημιουργικότητα μπορεί να αξιολογηθεί αντικειμενικά και ποσοτικά, αυτό δεν αρκεί ώστε ο επιστήμονας δεδομένων να κατασκευάσει ένα αυτόματο σύστημα αξιολόγησης δημιουργικής γραφής. Αυτό απαιτεί την εξαγωγή των χαρακτηριστικών κειμένου που σχετίζονται με τη δημιουργικότητα. Δυστυχώς, τα χαρακτηριστικά που καθιστούν ένα κείμενο δημιουργικό δεν ευθυγραμμίζονται με τα χαρακτηριστικά που το καθιστούν ευανάγνωστο [MCM10]:

- Ένα κείμενο μπορεί να χαρακτηρίζεται από υψηλή συνοχή, με μεγάλο πλήθος από συνδετικές λέξεις και coreferences, αλλά χωρίς να έχει ποιοτικό περιεχόμενο.
- Ένα κείμενο μπορεί να φαίνεται κακογραμμένο εκ πρώτης όψεως, χωρίς ιδιαίτερη συνοχή, αλλά να περιέχει πλούσια νοήματα, που ένα ανθρώπινος αναγνώστης μπορεί να εξάγει.

Με άλλα λόγια, η δημιουργικότητα και η ποιότητα του γραπτού λόγου φαίνονται να σχετίζονται με χαρακτηριστικά κειμένου που είναι μη απτά, πράγμα που μάλλον ενισχύει την άποψη ότι η δημιουργικότητα δεν μπορεί να αξιολογηθεί αντικειμενικά. Βέβαια, αυτή η αντίληψη δεν περιορίζει την έρευνα προς αυτήν την κατεύθυνση. Πράγματι, έχουν κατασκευαστεί επιτυχημένα regression models που επιτυγχάνουν υψηλή συσχέτιση μεταξύ προβλέψεων δημιουργικότητας και αξιολογήσεων κριτών χρησιμοποιώντας απλά γλωσσολογικά χαρακτηριστικά [ZHK09].

3 Παρουσίαση dataset & στατιστικά στοιχεία

Σε αυτήν την ενότητα περιγράφεται το dataset που θα χρησιμοποιηθεί στην εργασία στην υποενότητα 3.1, ενώ στην υποενότητα 3.2 εξάγονται κάποια στατιστικά στοιχεία για την καλύτερη κατανόησή του.

3.1 Dataset παραμυθιών

Το dataset που χρησιμοποιείται σε αυτήν την εργασία αφορά τη συνηθισμένη διαδικασία συμπλήρωσης ιστορίας. Συγκεκριμένα, δόθηκαν σε παιδιά σχολικής ηλικίας κάποια παραμύθια γραμμένα από αναγνωρισμένους συγγραφείς, με το τέλος των ιστοριών να λείπει. Στη συνέχεια, ζητήθηκε από τα παιδιά να συμπληρώσουν τα παραμύθια. Για όλα τα παραμύθια δίνεται η ηλικιακή ομάδα του συγγραφέα. Συγκεκριμένα, έχουμε τις εξής ηλικιακές κατηγορίες:

- Κατηγορία A, που αντιστοιχεί στις ηλικίες 6 μέχρι 8. Αυτή η κατηγορία αφορά παιδιά στις πρώτες τάξεις του δημοτικού.
- Κατηγορία B, που αντιστοιχεί στις ηλικίες 9 μέχρι 10. Αυτή η κατηγορία αφορά παιδιά στις μεσαίες τάξεις του δημοτικού.
- Κατηγορία C, που αντιστοιχεί στις ηλικίες 11 μέχρι 12. Αυτή η κατηγορία αφορά παιδιά στις τελευταίες τάξεις του δημοτικού.
- Κατηγορία D, που αντιστοιχεί σε ηλικίες μεγαλύτερες από τα 12. Αυτή η κατηγορία αφορά μεγαλύτερα παιδιά γυμνασίου και λυκείου.

Για λίγα από τα παραμύθια των παιδιών ζητήθηκε από ειδικούς του χώρου, όπως για παράδειγμα δασκάλους και θεατρικούς συγγραφείς, να αξιολογήσουν το επίπεδο της δημιουργικότητας στους εξής άξονες:

1. Ποσότητα Ιδεών (Νοητική ευχέρεια)
2. Είδη Ιδεών (Νοητική ευλυγισία/ευελιξία)
3. Σπανιότητα Ιδεών (Πρωτοτυπία της σκέψης)
4. Εικόνα/Λεκτική εικονοποίηση
5. Ήχος/Λεκτική περιγραφή ήχου
6. Περιγραφή χαρακτήρα/Χαρακτηρισμός
7. Αφήγηση/εξιστόρηση

Αυτά τα κριτήρια στο εξής θα αναφέρονται ως Κριτήριο 1, 2, 3, 4, 5, 6, 7, στις περιπτώσεις όπου η οικονομία χώρου είναι προτεραιότητα. Τα παραμύθια αξιολογήθηκαν σε κλίμακα από 1 έως 4, όπου θα μπορούσαμε να ερμηνεύσουμε το 1 ως φτωχό και το 4 ως εξαιρετικό. Το συγκεκριμένο dataset περιέχει τα κείμενα όπως γράφηκαν από τα παιδιά, και άρα επιτρέπει την εξαγωγή συσχετίσεων μεταξύ των γλωσσολογικών χαρακτηριστικών του γραπτού λόγου, των απόψεων των κριτών και της ηλικίας των συγγραφέων.

Τίτλος Παραμυθιού	Συγγραφέας	Κατηγορία				Σύνολο
		A	B	C	D	
Το αστέρι το 'σκασε	Φωτεινή Κωνσταντοπούλου	12	21	12	4	49
Γίνεται Πάσχα χωρίς κόκκινα αυγά;	Βαγγέλης Ηλιόπουλος	26	34	21	9	90
Τα χριστουγεννιάτικα κουλουράκια	Μαρία Ε. Μωραΐτου	36	58	55	12	161
Το μυστήριο του πασχαλινού λαγού	Κώστας Στοφόρος	59	64	42	8	173
Οι τρεις μάγοι στην Αθήνα	Πηγή Γρύλλη	31	22	17	11	81
Το πρώτο Πάσχα της ζωής μου	Κάρμεν Ρουγγέρι, Χριστίνα Κουλουμπή	80	73	69	20	242
Η πόλη που δεν θα γιόρταζε τα Χριστούγεννα	Ουρανία Μπούρτζινου	47	93	47	12	199

Πίνακας 1: Τίτλοι παραμυθιών, στοιχεία συγγραφέων, και πλήθος παιδιών που συμπλήρωσαν τα ημιτελή παραμύθια.

3.2 Στατιστική ανάλυση

Σε αυτήν την υποενότητα αναφέρονται στατιστικά στοιχεία του dataset των παραμυθιών, με στόχο την καλύτερη κατανόηση του είδους των δεδομένων. Στο dataset περιέχονται 7 διαφορετικά παραμύθια, και για καθένα από αυτά έχουμε αρκετές δεκάδες δημιουργικές συνεισφορές παιδιών για τον τρόπο με τον οποίο τελειώνουν. Στον Πίνακα 1 φαίνονται τα στοιχεία που αφορούν τα παραμύθια, καθώς και τα πλήθη των παιδιών που συνεισέφεραν τις ιστορίες τους.

Δυστυχώς, λόγω της δυσκολίας της διαδικασίας αξιολόγησης δημιουργικότητας, το dataset περιέχει αξιολογήσεις ειδικών μόνο για 22 ιστορίες παιδιών που συμπληρώνουν το παραμύθι «Το πρώτο Πάσχα της ζωής μου». Οι αξιολογήσεις αυτές αφορούν παιδιά των ηλικιακών κατηγοριών A, B, C, και η ανάλυσή μας θα περιοριστεί σε αυτές τις κατηγορίες στη συνέχεια της εργασίας. Στον Πίνακα 2 παρουσιάζονται οι μέσες αξιολογήσεις των κριτών για την κάθε ηλικιακή ομάδα. Παρατηρούμε ότι υπάρχουν μικρές μόνο διαφορές στον τρόπο με τον οποίο ο μέσος κριτής αξιολογεί τα κείμενα διαφορετικών ηλικιακών κατηγοριών. Αυτό που δεν αποκαλύπτει ο συγκεκριμένος μέσος όρος είναι οι έντονες διαφωνίες που εμφανίζονται μεταξύ των κριτών όσον αφορά κάποια κριτήρια. Στον Πίνακα 3 παρουσιάζονται τα κριτήρια δημιουργικότητας που οδήγησαν στα υψηλότερα και στα χαμηλότερα επίπεδα συμφωνίας μεταξύ των κριτών, με χρήση της μετρικής Pearson correlation. Αυτό μας δίνει μια αίσθηση του πόσο αντικειμενικά μπορούν να αξιολογηθούν οι αντίστοιχες εκφάνσεις της δημιουργικότητας και κατά πόσο πρέπει να χρησιμοποιηθούν σε μελλοντικά πειράματα.

Κατηγορία	Πλήθος	Κρ. 1	Κρ. 2	Κρ. 3	Κρ. 4	Κρ. 5	Κρ. 6	Κρ. 7
A	8	2.65	2.375	2.175	2.85	2.125	2.1	2.825
B	7	2.771	2.657	2.657	3.142	2.114	2.428	2.685
C	7	2.714	2.6	2.4	3.028	2.257	2.628	2.857

Πίνακας 2: Στατιστικά στοιχεία αξιολόγησης ειδικών για 22 ιστορίες που συμπληρώνουν το παραμύθι «Το πρώτο Πάσχα της ζωής μου». Για κάθε ηλικιακή ομάδα, εξάγουμε το μέσο όρο των αξιολογήσεων των κριτών για κάθε κριτήριο.

		<i>j</i>			
		2	3	4	5
<i>i</i>	1	0.666	0.650	0.638	0.735
	2		0.743	0.618	0.741
	3			0.612	0.769
	4				0.592

(α') Τιμές Pearson correlation για τα ζεύγη κριτών (i, j) στο κριτήριο λεκτικής περιγραφής ήχου (κριτήριο 5). Η μέση συμφωνία (μέσο Pearson correlation) είναι 0.676.

		<i>j</i>			
		2	3	4	5
<i>i</i>	1	0.437	0.402	-0.043	0.477
	2		0.193	0.248	0.458
	3			-0.073	0.481
	4				-0.091

(β') Τιμές Pearson correlation για τα ζεύγη κριτών (i, j) στο κριτήριο περιγραφής χαρακτήρα (κριτήριο 6). Η μέση συμφωνία (μέσο Pearson correlation) είναι 0.249.

Πίνακας 3: Παρατηρούμε ότι η μέγιστη συμφωνία μεταξύ των κριτών επιτυγχάνεται όσον αφορά το κριτήριο του ήχου/λεκτικής περιγραφής ήχου, ενώ η ελάχιστη συμφωνία παρατηρείται στο κριτήριο περιγραφής χαρακτήρα.

4 Εξαγωγή χαρακτηριστικών & εκπαίδευση μοντέλων

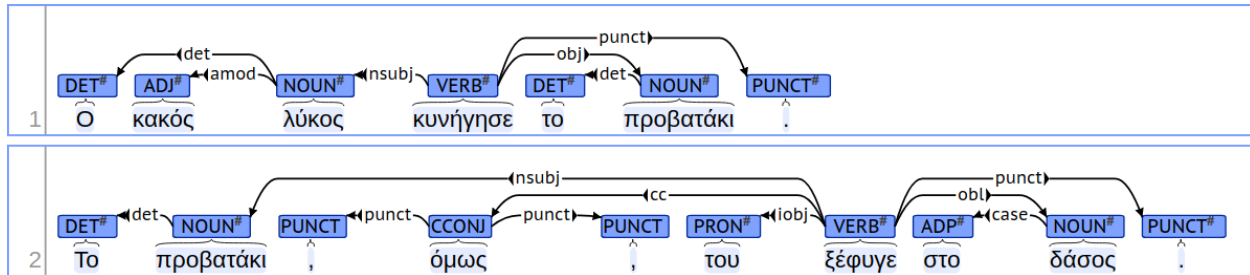
Στην ενότητα 3 περιγράφηκε το dataset των παραμυθιών όσον αφορά τον τρόπο κατασκευής του, καθώς και τους περιορισμούς που επιβάλλει όσον αφορά το πλήθος των διαθέσιμων αξιολογήσεων κριτών για τα επίπεδα δημιουργικότητας. Ωστόσο, το dataset αυτό δίνει την ευκαιρία να μελετήσουμε τον τρόπο με τον οποίο σχετίζονται γλωσσολογικά χαρακτηριστικά των κειμένων παιδιών με την ηλικία τους και τις αξιολογήσεις των ειδικών. Στην υποενότητα 4.1 περιγράφεται η εξαγωγή γλωσσολογικών χαρακτηριστικών με χρήση των υπηρεσιών του Ινστιτούτου Ελληνικού Λόγου (ΙΕΛ). Έπειτα, στην υποενότητα 4.2 μοντελοποιούνται με μοντέλα μηχανικής μάθησης οι εξαρτήσεις μεταξύ των γλωσσολογικών χαρακτηριστικών, της ηλικίας των συγγραφέων, και των αξιολογήσεων των κριτών.

4.1 Εξαγωγή χαρακτηριστικών

Για να μοντελοποιήσουμε τη δημιουργικότητα, θα πρέπει να συλλάβουμε την άρρητη γνώση που περιέχεται σε ένα κείμενο. Αυτό αποτελεί εξαιρετικά δύσκολο έργο, που δεν θα προσπαθήσουμε να προσεγγίσουμε. Ωστόσο, μπορούμε να μελετήσουμε τη χρησιμότητα «ρηχών» χαρακτηριστικών των κειμένων. Συγκεκριμένα, χρησιμοποιούμε τη διαδικτυακή υπηρεσία `ilsp_nlp_depparse_ud` του ΙΕΛ. Αυτή η υπηρεσία μας επιτρέπει να αναλύσουμε κείμενο γραμμένο στη φυσική γλώσσα, με στόχο να εξάγουμε όρους, μέρη του λόγου, λήμματα, αλλά και εξαρτήσεις μεταξύ όρων [PP17]. Για παράδειγμα, το απλό κείμενο «Ο κακός λύκος κυνήγησε το προβατάκι. Το προβατάκι, όμως, του ξέφυγε στο δάσος.» αναλύεται ως εξής:

```
# sent_id = o_outputout1
# text = 0 κακός λύκος κυνήγησε το προβατάκι.
1 0 ο DET AtDf Case=Nom|Definite=Def|Gender=Masc|Number=Sing|PronType=Art 3 det _ _
2 κακός κακός ADJ Aj Case=Nom|Gender=Masc|Number=Sing 3 amod _ _
3 λύκος λύκος NOUN NoCm Case=Nom|Gender=Masc|Number=Sing 4 nsubj _ _
4 κυνήγησε κυνηγώ VERB VbMn Aspect=Perf|Mood=Ind|Number=Sing|Person=3|Tense=Past|Voice=Act 0 root _ _
5 το ο DET AtDf Case=Acc|Definite=Def|Gender=Neut|Number=Sing|PronType=Art 6 det _ _
6 προβατάκι προβατάκι NOUN NoCm Case=Acc|Gender=Neut|Number=Sing 4 obj _ SpaceAfter=No
7 . . PUNCT PTERM_P PunctType=PTERMP 4 punct _ _
# sent_id = o_outputout2
# text = Το προβατάκι, όμως, του ξέφυγε στο δάσος.
1 Το ο DET AtDf Case=Nom|Definite=Def|Gender=Neut|Number=Sing|PronType=Art 2 det _ _
2 προβατάκι προβατάκι NOUN NoCm Case=Nom|Gender=Neut|Number=Sing 7 nsubj _ SpaceAfter=No
3 , , PUNCT PUNCT _ 4 punct _ _
4 όμως όμως CCONJ CjCo _ 7 cc _ SpaceAfter=No
5 , , PUNCT PUNCT _ 4 punct _ _
6 του εγώ PRON PnP Case=Gen|Gender=Masc|Number=Sing|Person=3|PersPronType=We|PronType=Prs 7 iobj _ _
7 ξέφυγε ξεφεύγω VERB VbMn Aspect=Perf|Mood=Ind|Number=Sing|Person=3|Tense=Past|Voice=Act 0 root _ _
8 στο του ADP AsPpPa Case=Acc|Gender=Neut|Number=Sing 9 case _ _
9 δάσος δάσος NOUN NoCm Case=Acc|Gender=Neut|Number=Sing 7 obl _ SpaceAfter=No
10 . . PUNCT PTERM_P PunctType=PTERMP 7 punct _ _
```

Το αντίστοιχο δέντρο εξαρτήσεων φαίνεται στην Εικόνα 1. Αυτή η υπηρεσία χρησιμοποιείται για να εξάγουμε τα εξής χαρακτηριστικά από τις καταλήξεις των παραμυθιών που έγραψαν τα παιδιά:



Εικόνα 1: Συντακτικό δέντρο που προέκυψε από την υπηρεσία του ΙΕΛ για το κείμενο «Ο κακός λύκος κυνήγησε το προβατάκι. Το προβατάκι, όμως, του ξέφυγε στο δάσος.».

- Πλήθος όρων που περιέχονται στο κείμενο. Υπερβολικά μικρά κείμενα ενδέχεται να μην είναι δημιουργικά ή να υποδεικνύουν συγγραφείς μικρής ηλικίας.
- Μέσο μήκος πρότασης. Μικρές προτάσεις μπορεί να υποδεικνύουν φτωχό λόγο ή μικρή ηλικία.
- Μέση απόσταση κύριου όρου (συνήθως ρήματος) από την αρχή της πρότασης. Μεγάλες αποστάσεις μπορεί να υποδεικνύουν μακριές εξαρτήσεις.
- Πλήθος όρων που εμφανίζονται μόνο στο συγκεκριμένο κείμενο και όχι σε κείμενα που έγραψαν άλλα παιδιά για να συμπληρώσουν το ίδιο παραμύθι. Η ύπαρξη τέτοιων όρων είναι μάλλον ένδειξη δημιουργικότητας.
- Ποσοστό επιθέτων, ουσιαστικών, και ρημάτων στο κείμενο. Το αν κάποιος χρησιμοποιεί πολλά επίθετα ή ουσιαστικά μπορεί να χαρακτηρίζει το είδος της γραφής του.
- Ποσοστό επιθέτων, ουσιαστικών, και ρημάτων που εμφανίζονται μόνο μια φορά στο κείμενο. Η μη επανάληψη μπορεί να χαρακτηρίζει συγγραφική ευχέρεια.
- Ποσοστό επιθέτων, ουσιαστικών, και ρημάτων που δεν εμφανίζονται στο κείμενο που τα παιδιά ζητήθηκαν να συμπληρώσουν. Αυτό χαρακτηρίζει το πόσο αποστασιοποιήθηκαν τα παιδιά από τους κύριους όρους που εμφανίζονται στο κείμενο που τους δόθηκε.

Συνολικά, έχουμε 13 γλωσσολογικά χαρακτηριστικά που εξάγονται από τα κείμενα. Παραπάνω περιγράφηκε ο τρόπος με τον οποίο αναμένεται να μας βοηθήσουν τα χαρακτηριστικά αυτά όσον αφορά την πρόβλεψη των επιπέδων δημιουργικότητας και της ηλικίας των συγγραφέων. Ωστόσο, το κατά πόσο αυτά τα χαρακτηριστικά είναι χρήσιμα θα φανεί μόνο αφού εκπαιδύσουμε μοντέλα που προσπαθούν να κάνουν αυτές τις προβλέψεις στην υποενότητα 4.2.

4.2 Εκπαίδευση μοντέλων μηχανικής μάθησης

Εδώ μελετούμε τις εξής συσχετίσεις:

1. Με βάση τα γλωσσολογικά χαρακτηριστικά των κειμένων, πώς μπορούμε να προβλέψουμε τη μέση αξιολόγηση των κριτών για τα κείμενα αυτά;

2. Με βάση τις αξιολογήσεις των κριτών, πώς μπορούμε να προβλέψουμε τις ηλικιακές ομάδες των συγγραφέων;
3. Με βάση τα γλωσσολογικά χαρακτηριστικά των κειμένων, πώς μπορούμε να προβλέψουμε τις ηλικιακές ομάδες των συγγραφέων;

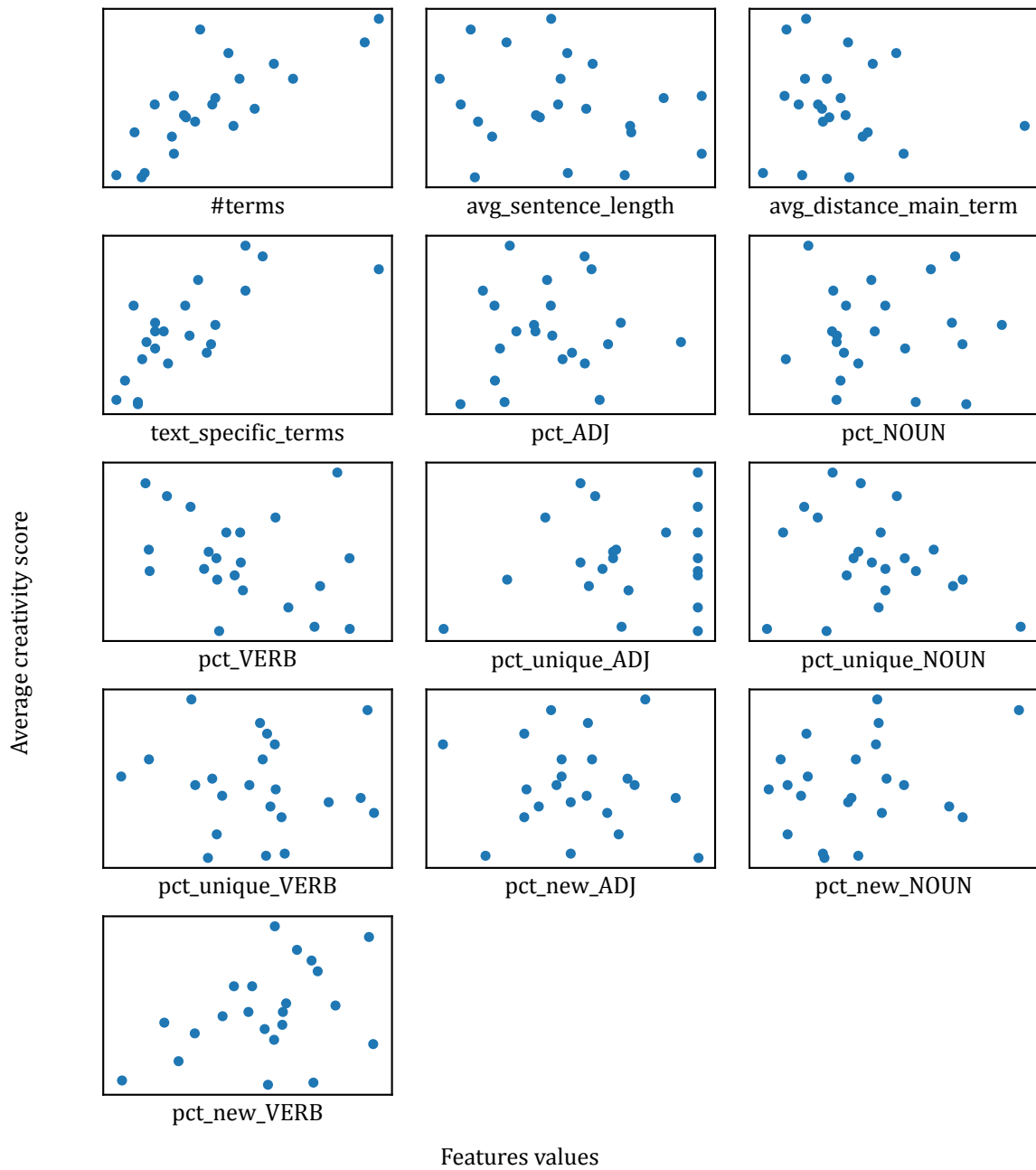
Λόγω των περιορισμών που επιβάλλει το dataset, μπορούμε να χρησιμοποιήσουμε το σύνολο των διαθέσιμων κειμένων μόνο για τη μελέτη της τρίτης συσχέτισης, καθώς για τις πρώτες δύο έχουμε διαθέσιμα μόνο 22 παραμύθια με αξιολογήσεις κριτών. Επίσης, όπως σημειώνεται παραπάνω, η ανάλυσή μας θα περιοριστεί στις ηλικιακές κατηγορίες, A, B και C, και όχι στους συγγραφείς μεγαλύτερης ηλικίας. Τέλος, για τις ανάγκες της ανάλυσης δεδομένων και κατασκευής μοντέλων πρόβλεψης έχει χρησιμοποιηθεί σε όλη την έκταση της εργασίας η βιβλιοθήκη scikit-learn της Python.

4.2.1 Πρόβλεψη αξιολογήσεων κριτών με βάση γλωσσολογικά χαρακτηριστικά

Χρησιμοποιούμε τα 13 γλωσσολογικά χαρακτηριστικά που περιγράφηκαν στην υποενότητα 4.1 με στόχο να κατασκευάσουμε regression models που να προβλέπουν τις μέσες αξιολογήσεις των κριτών για το καθένα από τα 7 κριτήρια δημιουργικότητας που περιγράφηκαν στην υποενότητα 3.1. Συγκεκριμένα, τα κριτήρια αντιμετωπίζονται ξεχωριστά, και ακολουθούνται τα εξής βήματα:

1. Για κάθε κείμενο, εξάγουμε τον μέσο όρο των αξιολογήσεων των κριτών για το κριτήριο που μας ενδιαφέρει.
2. Για ένα πλήθος μοντέλων μηχανικής μάθησης, που περιλαμβάνει νευρωνικά δίκτυα, SVMs, ελάχιστα τετράγωνα, και άλλα, εφαρμόζουμε την τεχνική leave-one-out, δηλαδή για το κάθε κείμενο επιλέγουμε να το «αφήσουμε εκτός», εκπαιδεύουμε το μοντέλο με τα υπόλοιπα κείμενα, και κάνουμε την πρόβλεψή μας στο δείγμα που αφήσαμε εκτός της εκπαίδευσης. Αυτό γίνεται ώστε να εξάγουμε όσο το δυνατόν περισσότερη πληροφορία από τα περιορισμένα παραδείγματα εκπαίδευσης που έχουμε.

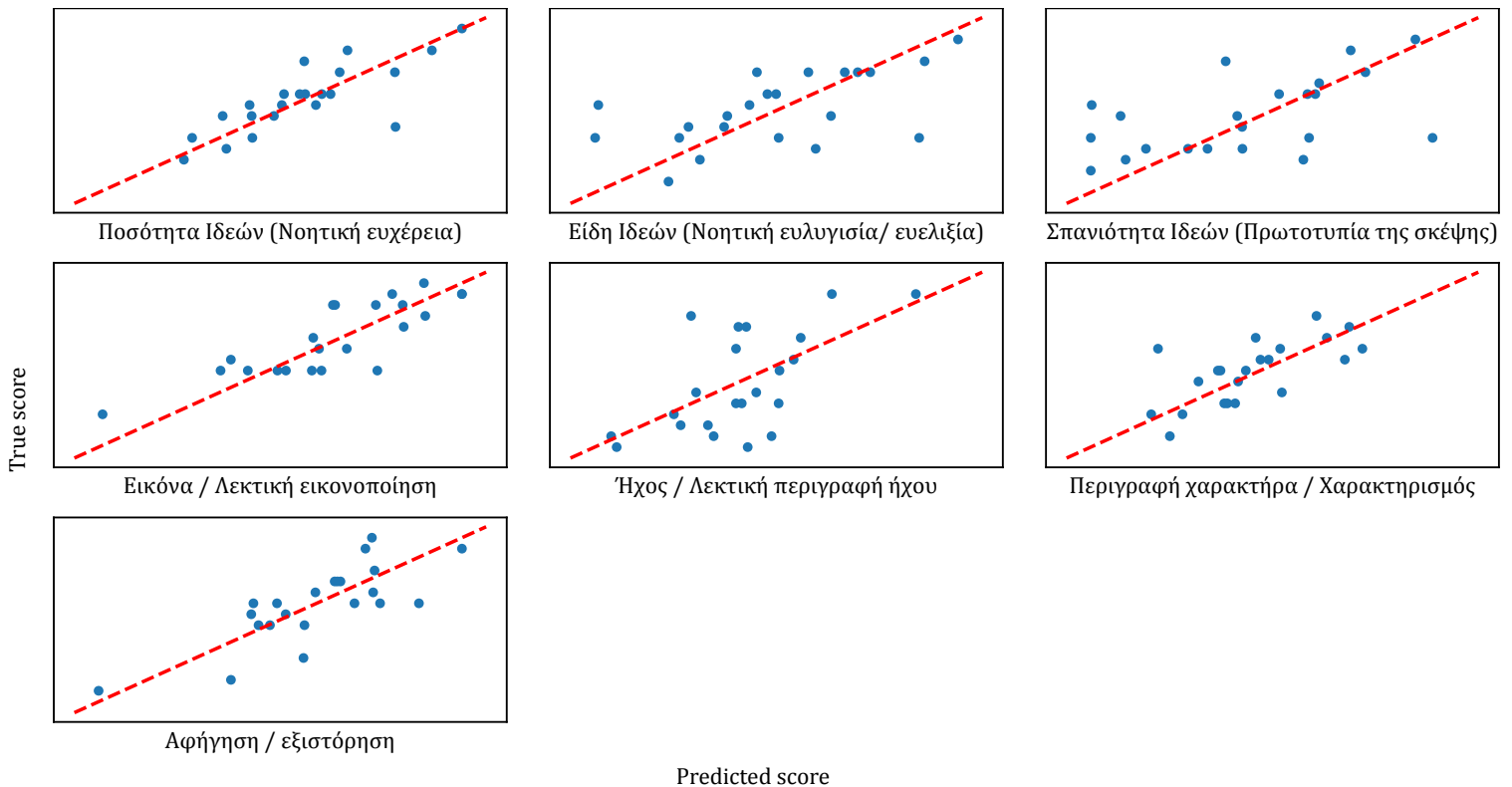
Στον Πίνακα 4 παρουσιάζονται τα αποτελέσματα αυτής της μοντελοποίησης. Για λόγους σύγκρισης, παρατίθεται και το μέσο correlation των κριτών με τη μέση τιμή. Παρατηρούμε ότι διαφορετικά μοντέλα μηχανικής μάθησης αποδίδουν καλύτερα για την πρόβλεψη διαφορετικών κριτηρίων δημιουργικότητας, αλλά η ακρίβεια του αποτελέσματος είναι ικανοποιητική. Μάλιστα, βλέπουμε ότι για ορισμένα κριτήρια, όπως η περιγραφή χαρακτήρα, το μοντέλο που εκπαιδεύσαμε παρουσιάζει καλύτερο correlation με τη μέση αξιολόγηση σε σχέση με τους κριτές. Επιπλέον, βλέπουμε ότι τα ίδια μοντέλα φαίνεται να αποδίδουν καλύτερα σε όλα τα κριτήρια, και αυτά τα μοντέλα είναι τα SVR (Support Vector Regression), και Linear Regression (ελάχιστα τετράγωνα). Εξάριση αποτελεί το μοντέλο Ridge Regression CV (ελάχιστα τετράγωνα με regularization & cross-validation), που φαίνεται να δίνει καλύτερα αποτελέσματα στο κριτήριο της λεκτικής περιγραφής ήχου. Για να φανεί η επίδραση του κάθε χαρακτηριστικού στην κρίση των ειδικών, παρουσιάζεται ο τρόπος με τον οποίο μεταβάλλεται η μέση αξιολόγηση συναρτήσει των χαρακτηριστικών στην Εικόνα 2. Τέλος, παρουσιάζεται η σχέση των προβλεπόμενων από το μοντέλο και των πραγματικών μέσων τιμών για το κάθε κριτήριο στην Εικόνα 3.



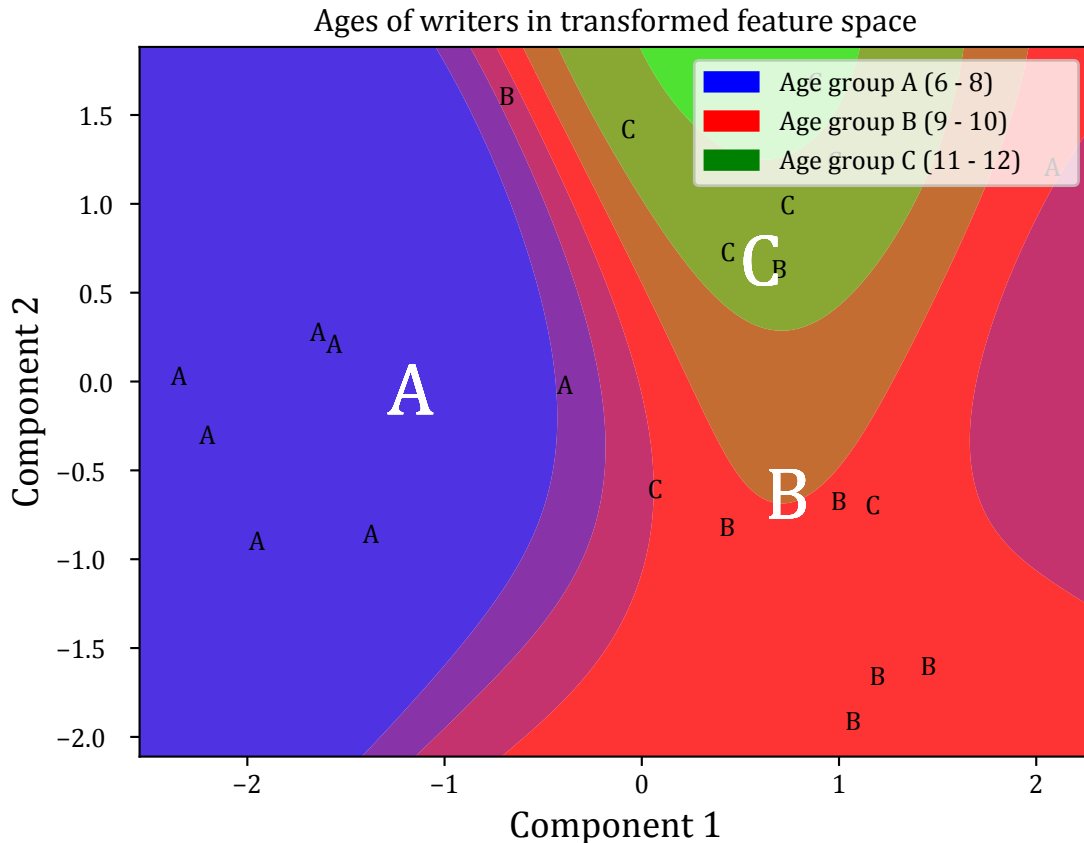
Εικόνα 2: Μεταβολή της μέσης αξιολόγησης κριτών συναρτήσει του καθενός εκ των 13 χαρακτηριστικών.

Κριτήριο	Correlation Κριτών με Μ.Ο.	Correlation Μοντέλου με Μ.Ο.	Μοντέλο
Κρ. 1	0.784	0.799	SVR
Κρ. 2	0.707	0.589	Linear Regression
Κρ. 3	0.699	0.528	Linear Regression
Κρ. 4	0.789	0.829	Linear Regression
Κρ. 5	0.860	0.623	Ridge Regression CV
Κρ. 6	0.628	0.681	SVR
Κρ. 7	0.773	0.779	SVR

Πίνακας 4: Σύγκριση καλύτερων μοντέλων πρόβλεψης αξιολογήσεων κριτών με τις ίδιες τις αξιολογήσεις, όσον αφορά το correlation με τον μέσο όρο.



Εικόνα 3: Μεταβολή της μέσης αξιολόγησης κριτών συναρτήσει του καθενός εκ των 13 χαρακτηριστικών.



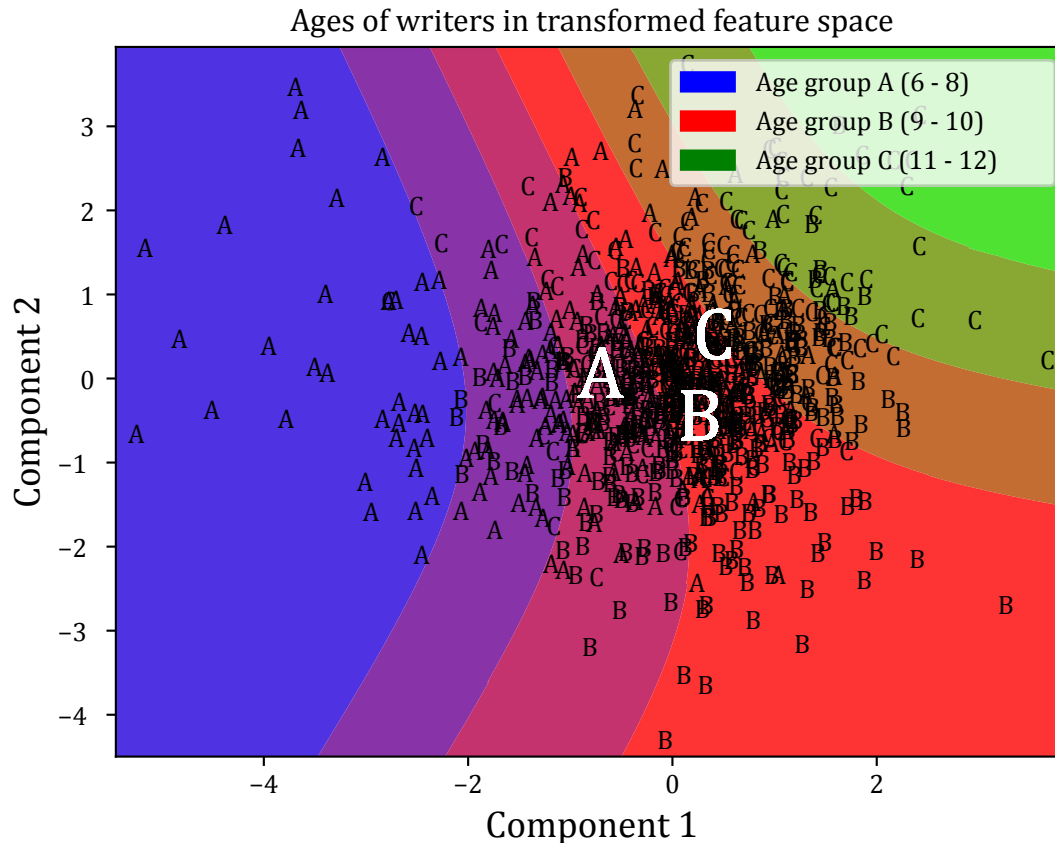
Εικόνα 4: Παρουσίαση των κειμένων στον μετασχηματισμένο χώρο χαρακτηριστικών αξιολογήσεων κριτών. Τα επιμέρους κείμενα εμφανίζονται με μικρά γράμματα, ενώ με μεγάλα φαίνονται οι μέσες τιμές των κλάσεων στον χώρο των χαρακτηριστικών.

4.2.2 Πρόβλεψη ηλικιακών ομάδων με βάση τις αξιολογήσεις των κριτών

Χρησιμοποιούμε τις αξιολογήσεις των κριτών ως χαρακτηριστικά για να προβλέψουμε τις ηλικιακές ομάδες των συγγραφέων. Κάποια στατιστικά που αφορούν αυτή τη συσχέτιση έχουν παρουσιαστεί ήδη στον Πίνακα 2. Έτσι, για κάθε κείμενο έχουμε $7 \text{ κριτήρια} \times 5 \text{ κριτές} = 35$ χαρακτηριστικά και μια ετικέτα (label), που είναι A, B, ή C. Δηλαδή, έχουμε ένα απλό πρόβλημα classification. Τα χαρακτηριστικά εμπλουτίζονται με χρήση πολυωνυμικών συνδυασμών χαρακτηριστικών βαθμού 2. Έπειτα, για την καλύτερη οπτικοποίηση των δεδομένων χρησιμοποιούμε Linear Discriminant Analysis, ώστε να μειώσουμε τη διάσταση των δεδομένων και να κρατήσουμε μόνο 2 χαρακτηριστικά. Το αποτέλεσμα αυτής της ανάλυσης φαίνεται στην Εικόνα 4. Το καλύτερο μοντέλο για την πρόβλεψη των ηλικιακών ομάδων με βάση τις αξιολογήσεις των ειδικών φαίνεται να είναι το Naive Bayes με υπόθεση κανονικής κατανομής, που δίνει ακρίβεια 71.67% με χρήση 2-fold cross-validation.

4.2.3 Πρόβλεψη ηλικιακών ομάδων με βάση γλωσσολογικά χαρακτηριστικά

Χρησιμοποιούμε τα γλωσσολογικά χαρακτηριστικά όλων των διαθέσιμων κειμένων, με στόχο να προβλέψουμε την ηλικία των συγγραφέων. Λόγω των περιορισμών που επιβάλ-



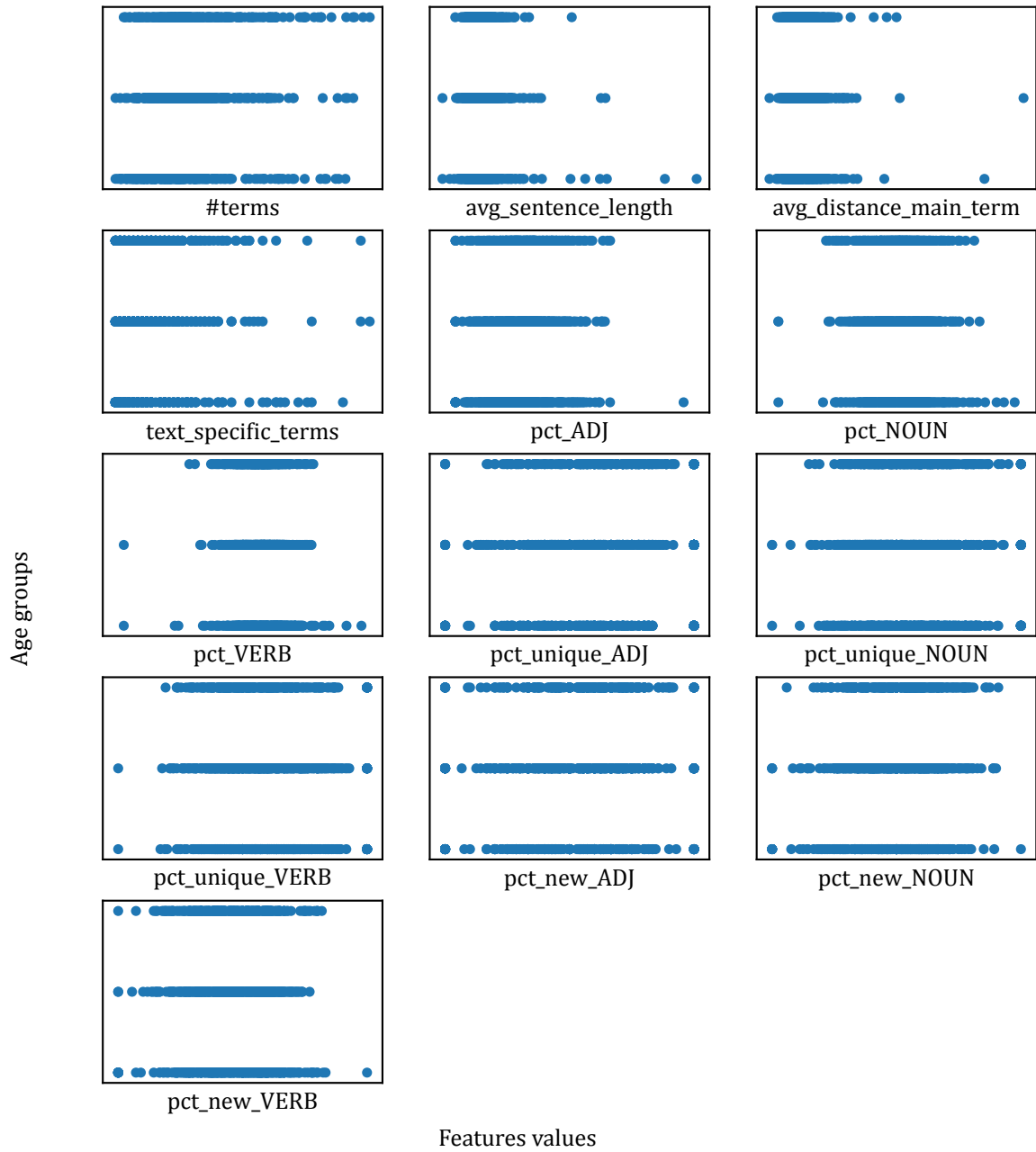
Εικόνα 5: Παρουσίαση των κειμένων στον μετασχηματισμένο χώρο χαρακτηριστικών γλωσσολογικών χαρακτηριστικών και των περιοχών απόφασης του μοντέλου linear SVM. Τα επιμέρους κείμενα εμφανίζονται με μικρά γράμματα, ενώ με μεγάλα φαίνονται οι μέσες τιμές των κλάσεων στον χώρο των χαρακτηριστικών.

λουν οι υπηρεσίες του ΙΕΛ, κάποια κείμενα έμειναν εκτός της ανάλυσής μας. Το τελικό dataset αποτελείται από περίπου 1000 παραμύθια. Το μοντέλο που αποδίδει καλύτερα σε μια διαδικασία 10-fold cross-validation είναι το SVM με linear kernel, που δίνει ακρίβεια πρόβλεψης 55.24%. Χρησιμοποιώντας ξανά Linear Discriminant Analysis λαμβάνουμε την Εικόνα 5, όπου φαίνονται τα κείμενα στον μετασχηματισμένο χώρο χαρακτηριστικών, καθώς και οι περιοχές απόφασης του linear SVM classifier. Για να φανεί η επίδραση του κάθε χαρακτηριστικού στην κρίση των ειδικών, παρουσιάζεται ο τρόπος με τον οποίο μεταβάλλεται η μέση αξιολόγηση συναρτήσει των χαρακτηριστικών στην Εικόνα 6.

5 Συμπεράσματα

Σε αυτήν την εργασία ερευνήθηκαν οι σχέσεις μεταξύ γλωσσολογικών χαρακτηριστικών, ηλικίας συγγραφέων και αξιολογήσεων κριτών για συνέχειες παραμυθιών γραμμένες από παιδιά δημοτικού. Τα βασικά συμπεράσματα που εξάγουμε είναι:

- Η αξιολόγηση της δημιουργικότητας από τους κριτές είναι σε μεγάλο βαθμό υποκειμενική, όπως φαίνεται από τα στατιστικά στοιχεία που εξάχθηκαν στην ενότητα 3.



Εικόνα 6: Μεταβολή της μέσης αξιολόγησης κριτών συναρτήσει του καθενός εκ των 13 χαρακτηριστικών.

- Τα απλά χαρακτηριστικά που χρησιμοποιήθηκαν στην ενότητα 4 αρκούν για να κατασκευάσουμε ένα regression model που να συμφωνεί σε ικανοποιητικό βαθμό με τη μέση ανθρώπινη κρίση για τη δημιουργικότητα. Επίσης, η συνεισφορά των χαρακτηριστικών αποκαλύπτει ότι κάποια πολύ βασικά κριτήρια, όπως το πλήθος των όρων που εμφανίζονται σε ένα κείμενο, επηρεάζουν ουσιαστικά τη μέση αξιολόγηση των κριτών.
- Η πρόβλεψη της ηλικιακής κατηγορίας φαίνεται να γίνεται πιο εύκολα χρησιμοποιώντας τις αξιολογήσεις των κριτών παρά τα γλωσσολογικά χαρακτηριστικά των κειμένων.

Τονίζεται ότι τα παραπάνω συμπεράσματα είναι αρκετά επισφαλή, ειδικά όσον αφορά το κομμάτι της δημιουργικότητας, λόγω του μικρού πλήθους παραμυθιών για τα οποία είναι διαθέσιμη αξιολόγηση ειδικών.

Αναφορές

- [DSR15] Saray Díaz Suarez and Alan Reeves. Evaluating creative writing. 2015.
- [MCM10] Danielle S. McNamara, Scott A. Crossley, and Philip M. McCarthy. Linguistic features of writing quality. *Written Communication*, 27(1):57–86, 2010. URL: <https://doi.org/10.1177/0741088309351547>, arXiv:<https://doi.org/10.1177/0741088309351547>, doi:10.1177/0741088309351547.
- [Moz13] Hamideh Mozaffari. An analytical rubric for assessing creativity in creative writing. 3, 12 2013.
- [PP17] Prokopis Prokopidis and Haris Papageorgiou. Universal dependencies for greek. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 102–106, 2017.
- [ZXK09] Xiaojin Zhu, Zhiting Xu, and Tushar Khot. How creative is your writing? a linguistic creativity measure from computer science and cognitive psychology perspectives. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity, CALC '09*, pages 87–93, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL: <http://dl.acm.org/citation.cfm?id=1642011.1642023>.